04-1465
(Serial No. 09/619,643)

UNITED STATES COURT OF APPEALS FOR THE FEDERAL CIRCUIT

IN RE DANE K. FISHER AND RAGHUNATH V. LALGUDI

Appeal from the United States Patent and Trademark Office Board of Patent Appeals and Interferences in Appeal No. 2002-2046.

BRIEF FOR APPELLANTS

DANE K. FISCHER AND RAGHUNATH V. LALGUDI

Seth P. Waxman
Wilmer Cutler Pickering
Hale and Dorr LLP
2445 M Street, N.W.
Washington, D.C. 20037
(202) 663-6800

William F. Lee
Richard W. O'Neill
Wilmer Cutler Pickering
Hale and Dorr LLP
60 State Street
Boston, MA 02109
(617) 526-6000

William G. McElwain
Henry N. Wixon
Wilmer Cutler Pickering
Hale and Dorr LLP
1455 Pennsylvania Ave., N.W.
Washington, D.C. 20004
(202) 942-8406

September 27, 2004

CERTIFICATE OF INTEREST

Counsel for Appellants, Dane K. Fischer and Raghunath V. Lalgudi, certifies the following:

1. The full name of every party or amicus represented by me is:

   Dane K. Fisher
   Raghunath V. Lalgudi

2. The name of the real party in interest represented by me is:

   Monsanto Technology LLC

3. All parent corporations and any publicly held companies that own 10% or more of the stock of the party represented by me are:

   Monsanto Company owns 100% of Monsanto Technology LLC.

4. The names of all law firms and the partners or associates that appeared for the party or amicus now represented by me before the trial court or agency, or expected to appear in this court, are:

   WILMER CUTLER PICKERING HALE AND DORR LLP

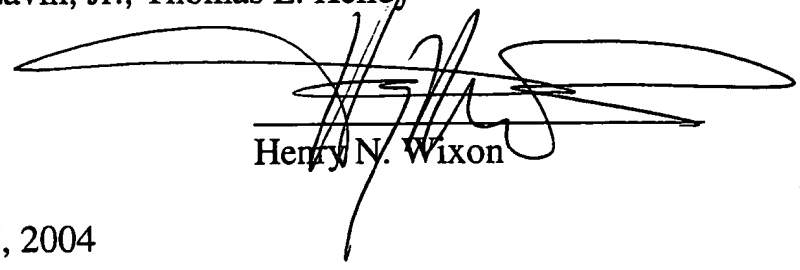
   Seth P. Waxman; William F. Lee; William G. McElwain; Henry N. Wixon; Richard W. O'Neill

   ARNOLD & PORTER

   David R. Marsh; Thomas E. Holsten; Holly L. Prutz; June E. Cohan

   MONSANTO COMPANY

   Lawrence M. Lavin, Jr.; Thomas E. Kelley



Henry N. Wixon

Dated: September 27, 2004

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

Constitutions and Statutes

Other Authority

STATEMENT OF RELATED CASES

There is no other appeal in or from this proceeding that was previously before this or any other appellate court. Notices of appeal have been filed in connection with the rejection of U.S. Patent Application Serial Nos. 09/654,617 (*In re Kovalic*), 09/620,392 (*In re Boukharov*), 09/540,232 (*In re Andersen*), 09/440,687 (*In re Byrum*), 09/565,240 (*In re Adab*), and 09/540,215 (*In re Lalgudi*). These cases, which have not yet been docketed for appeal, may be directly affected by this Court's decision in the pending appeal.

JURISDICTIONAL STATEMENT

The U.S. Patent and Trademark Office Board of Patent Appeals and Interferences ("the Board") had subject matter jurisdiction over this proceeding pursuant to 35 U.S.C. § 134. On March 31, 2004, the Board entered its Decision on Appeal ("Decision") affirming the Examiner's final rejection of claim 1 of U.S. Patent Application No. 09/619,643 ("the '643 Application'"). (JA0001-27.)[1] On May 27, 2004, the Applicants filed a timely Notice of Appeal from the Board's Decision in accordance with 35 U.S.C. §§ 141 and 142. (JA0260.) The Court has jurisdiction over this appeal under 28 U.S.C. § 1295(a)(4) and 35 U.S.C. §§ 141 and 144.

---

[1] The prefix "JA" refers to the Joint Appendix to be submitted by the parties.

I. STATEMENT OF ISSUES

1. Whether the Board erred by concluding that an expressed sequence tag (EST) is subject to a heightened standard of utility under 35 U.S.C. § 101 that hinges upon some undefined "spectrum" of knowledge about the function of the gene that corresponds to the EST?

2. Whether the Board erred by concluding that ESTs corresponding to genes of unknown function are incapable of satisfying the utility requirement of 35 U.S.C. § 101, even though ***all*** ESTs, including each of the claimed ESTs, can be used as research tools to provide one or more specific, substantial, and commercially valuable benefits to the scientific community?

II. STATEMENT OF THE CASE

On July 19, 2000, Dane K. Fischer and Raghunath V. Lalgudi (collectively, "Applicants") filed the '643 Application with the U.S. Patent and Trademark Office (PTO). (JA0028-115.) The '643 Application, entitled "Nucleic Acid Molecules and Other Molecules Associated with Plants," contains a single claim

directed to five purified nucleic acid molecules that code proteins, or fragments of proteins, found in maize plants. (JA0114; JA0169.)[2]

The Examiner issued a final rejection of claim 1 of the '643 Application on September 6, 2001, finding that the claim: (1) lacked utility under 35 U.S.C. § 101; (2) failed to satisfy the enablement and written description requirements of 35 U.S.C. § 112; and (3) was anticipated by two prior art references under 35 U.S.C. § 102. (JA0182-99.) The Applicants timely filed a Notice of Appeal from the Examiner to the Board of Patent Appeals and Interferences on December 6, 2001. (JA0201.)

The Examiner withdrew his anticipation rejection while this proceeding was pending on appeal before the Board. (JA0239-40.) In its Decision dated March 31, 2004, the Board reversed the Examiner's written description rejection. However, the Board affirmed the Examiner's final rejection of claim 1 for failure to satisfy the utility requirement of Section 101 and the enablement requirement of Section 112. (JA001-27.) This appeal from the Board's Decision followed.

---

[2] As initially filed, the '643 Application included seven claims directed to 4,013 nucleic acid molecule sequences. However, in January 2001, the Examiner issued a restriction requirement ordering the Applicants to elect certain claims and to limit their invention to "no more than five of the individual sequences for examination." (JA0146-50.) In response, the Applicants withdrew claims 2-7 and limited remaining claim 1 to five nucleic acid sequences (SEQ ID NO: 1 through SEQ ID NO: 5). (JA0153-54.)

III. STATEMENT OF FACTS

A. Relevant Principles of Molecular Genetics[3]

Each maize plant contains billions of cells that generate the enormous number of different proteins essential to the plant's proper growth, development, and function. For example, maize plants utilize proteins to create cell structure, process nutrients, generate new tissue, and fight disease and insects. Proteins also control the development and appearance of important physical plant traits such as color, flavor, sweetness, and stem and leaf size. (JA0057:21-JA0058:6.)

In recent decades, scientists have started to explore the complex genetic underpinnings of the intricate process used by maize plants to synthesize proteins. These efforts have resulted in a greater understanding of the maize genome, leading to the development of genetically altered maize plants that, for example, offer greater resistance to certain insects and herbicides. (JA0069:21-23.) Efforts to explore and further understand the maize genome continue today.

1. The Role of Chromosomes and Genes

The specific proteins produced by a maize plant cell are determined by the genetic code, or genome, of the plant. This genetic information is chemically

---

[3] This Court repeatedly has addressed the basic principles of molecular genetics. *See, e.g., In re Deuel*, 51 F.3d 1552, 1554-55 (Fed. Cir. 1995); *Amgen, Inc. v. Chugai Pharm. Co., Ltd.*, 927 F.2d 1200, 1207-08 (Fed. Cir. 1991); *In re O'Farrell*, 853 F.2d 894, 895-99 (Fed. Cir. 1988). For the convenience of the Court, the relevant principles from those cases are restated here.

stored within the nucleus of each plant cell in long supercoiled strands of deoxyribonucleic acid (DNA) called "chromosomes."

The portion of a chromosome that contains the genetic coding information necessary to make a particular protein is called a "gene." Structurally, genes are comprised of several components, including: (1) regulatory regions that affect the "expression" (i.e., synthesis) of a particular protein; (2) exons, which are the coding sequences of a gene that serve as the template for protein expression; and (3) introns, which are non-protein-coding sequences that exist between exons.[4]

2. The Role of DNA

DNA acts as the blueprint for all of the protein-driven activities that are necessary for a maize plant to develop, grow, and live. DNA molecules are comprised of repeating units called nucleotides that link together into long strands. The four nucleotides found in DNA – adenine (A), guanine (G), cystosine (C), and thymine (T) – are called bases, and the particular order of the linked nucleotide bases is referred to as the DNA sequence.

Each DNA molecule is comprised of two strands of nucleotide bases. The sequence of nucleotide bases found in one strand will "hybridize" (i.e., pair or bind) with the complementary sequence of nucleotide bases found in the other strand; adenine will hybridize with thymine (A-T), and guanine will hybridize with

---

[4] Each maize plant cell contains ten chromosomes that collectively encompass approximately 50,000 genes. (JA0421-22.)

cystosine (G-C). This hybridization of nucleotide bases results in a double-stranded DNA molecule that takes the form of a twisting double helix. Because of the unique hybridization properties of the four DNA nucleotide bases, the known sequence of one strand can be used to predict the complimentary sequence of the other strand.[5]

The DNA sequence of a gene contains all of the coding information necessary to produce a particular protein. However, the entire sequence is not translated directly into protein. Rather, only the protein-coding regions (the exons) of the gene are used as a template for protein synthesis. Within that protein-coding region, sequential groupings of three nucleotides ("codons") code for a single amino acid (the building blocks of proteins). The sequence of codons determines the amino acids in the resulting protein.

3. The Role of mRNA

Although the DNA sequence of a gene contains all of the information necessary to generate a particular protein, the DNA molecule itself is not directly involved in creation of the protein. Instead, when a gene is expressed, the relevant DNA sequence first is "transcribed" (i.e., copied) into a new single-strand of genetic material called messenger ribonucleic acid (mRNA). After the transcription of DNA into mRNA, the non-coding sequences (the introns) are

---

[5] The DNA found in each maize cell contains more than two billion nucleotide base pairs.

removed in a process called "splicing," leaving the coding sequence (the exons) necessary to produce the specific protein that corresponds to the gene. Once transcribed and spliced of introns, mRNA is transported outside of the cell nucleus and used to synthesize protein in a process called "translation;" the particular sequence of codons found in the mRNA are translated into a sequence of amino acids that comprise a protein.[6]

A cell generates mRNA only when a gene is being expressed. As such, scientists can determine that a particular gene is being expressed in certain tissue at a given point in time simply by confirming the existence of mRNA corresponding to the gene within the cells of the tissue.

4. The Role of cDNA

Because mRNA contains the same protein-coding regions (the exons) found in the DNA sequence from which it was derived, scientists can use mRNA as a tool to trace an expressed protein back to its originating gene. However, mRNA is quite unstable once extracted from a cell, making it a difficult object of study within a laboratory environment. Therefore, scientists typically use a process called "reverse transcription" (and a catalyst enzyme called "reverse transcriptase") to transcribe mRNA into a purified complementary DNA (cDNA) molecule

---

[6] Like DNA, mRNA is comprised of chains of four nucleotide bases. DNA and mRNA share three of the same bases (adenine, guanine, and cystosine), but mRNA contains uracil rather than thymine.

commonly referred to as a "clone." Like naturally occurring DNA, a man-made cDNA clone is comprised of two strands of nucleotide bases that take the form of a twisting double helix – the first strand is generated from (and thus, is "complementary" to) the single-stranded mRNA molecule, and the second strand is synthesized from the first strand of the clone.

A cDNA clone contains the same nucleotide sequence found in the mRNA from which it is generated; that is, the sequence derived from the exon portions of the corresponding gene. Using a variety of different sequencing processes, geneticists can determine the full or partial sequence of nucleotides forming a cDNA clone. Once the sequence of a cDNA clone is known, the codons found in that sequence can be used to determine the corresponding protein sequence. That information then can be used to determine the specific function of the protein expressed by the gene to which the cDNA clone corresponds.

5. The Role of cDNA Libraries

To study the specific genes being expressed in a specific tissue of an organism at a specific point in time, geneticists commonly construct a cDNA library of the tissue. These libraries take advantage of the fact that cells generate mRNA only when one or more genes within the cell are being expressed. Therefore, by extracting mRNA from the cells of specific tissue at a certain point in time (and using reverse transcription to convert mRNA into cDNA), a library of

cDNA clones for the tissue can be generated. By sequencing and studying the clones found in a tissue-specific cDNA library, scientists can determine which genes were being expressed in the tissue at the time of mRNA extraction.

6. The Role of Expressed Sequence Tags

To sequence the full length of every clone stored in a typical cDNA library (and other similar collections of genetic material) would present an extremely time consuming and costly endeavor. As such, geneticists have developed a set of research tools to help screen genetic libraries for genes and gene fragments of interest in a rapid and cost-effective manner. One such tool is an "expressed sequence tag" (EST) – a short nucleotide sequence (usually 150 to 450 nucleotides in length) that uniquely represents a fragment of a cDNA clone, and thus, a fragment of the protein-coding portion of an expressed gene. (JA0028:12-14; JA0031:7-9; JA0273-78.)

An EST typically is generated by isolating a random clone found in a cDNA library and then sequencing a small number of nucleotides from the end of one the clone's two strands. (JA0028:12-16; JA0030:13-JA0031:6.) When used as a probe and introduced into a sample containing a mixture of DNA (e.g., a cDNA library), a fragment of DNA corresponding to the EST sequence will hybridize under appropriate conditions with DNA molecules in the sample to which the EST uniquely corresponds. Successful hybridization confirms that the gene

corresponding to the EST was being expressed in the sample tissue at the time of mRNA extraction. (JA0031:7-8; JA0043:16-JA0044:10; JA0273-78.)[7]

Capitalizing upon this scientific property, geneticists routinely utilize ESTs to screen large cDNA libraries for the presence of specific expressed genes. (JA0031:7-8; JA0038:19-25; JA0273-78.) The information derived from these screens can be compiled into large digital databases and then analyzed with software tools in connection with a wide array of scientific applications, including, for example, activities such as genome mapping and linkage analysis. (JA0031:22-JA0035:26; JA0103:3-JA0105:11; JA0273-78.)

B. ESTs Have Significant Commercial Value in the Marketplace.

In the decade since their introduction, ESTs have helped revolutionize the field of genetics by providing scientists with a variety of powerful and commercially valuable research tools.[8] This importance is underscored by the

---

[7] An EST probe can emit a visible light with a distinguishable wavelength after it binds to a complementary target sequence. (JA0043:9-13.)

[8] Prior to the development of ESTs, scientists used short DNA fragments called "sequenced tagged sites" (STSs) to screen large populations of genetic material. STSs are not necessarily derived from expressed genes. Therefore, although STSs can be used like ESTs in some respects, ESTs are more useful research tools than anonymous STSs because ESTs represent genes that are known to express protein in a specific tissue at a certain point in time. In fact, the discovery of ESTs is widely credited for the completion of the Human Genome Project significantly ahead of schedule and well below budget. (JA0273-78.)

skyrocketing commercial industry that is premised on the usefulness of ESTs as genetic research tools.

The Applicants are not alone in their pursuit of ESTs. There is no dispute that numerous well-known biotechnology companies have dedicated substantial time, effort, and financial resources to research, discover, and utilize ESTs with respect to a variety of organisms, including maize plants. (JA0273-78.) Nor is there any dispute that many of these same companies have collectively derived hundreds of millions of dollars in revenues from licensing databases of ESTs that correspond to genes both of known ***and*** unknown function. (JA0336-416.)[9] For example:

- As of 1998, Pioneer Hi-Bred and DuPont claimed to have developed a combined EST database representing more than 75% of the maize genome (JA0382);
- In 1998, DuPont signed a five-year deal with Lynx Therapeutics to use Lynx's technology to organize Dupont's extensive crop EST databases and to provide genomic maps for crops in an effort to improve yield and agronomic traits such as drought tolerance (JA0377-78);
- Incyte Pharmaceuticals reported 1998 revenues from subscriptions to its EST databases at $105.6 million (JA0396-403);

---

[9] The existence of a substantial industry directed to the use of ESTs would have been well known to the Board. For that reason, although the Applicants raised issues of commercial success in their brief to the Board (JA0212), the record below did not include the specific references cited here. They are included in the Joint Appendix for the convenience of the Court.

- Celera Genomics has licensed its "Human Gene Index" EST database to companies such as Amgen, Inc., Novartis, and Pharmacia & Upjohn to, among other things, "enable and accelerate ... [the] identification of novel genes and factors that regulate and control gene expression" (JA0336-38; JA0341-46);

- Gene Logic, Inc. licenses its EST databases to numerous companies such as Procter & Gamble Pharmaceuticals and Organin, N.V., and in 1999 entered into an agreement with Affymetrix, Inc. to build a large commercial EST database for drug development (JA0349-51; JA0355; JA0406-09);

- Exelixis Pharmaceuticals Inc. and Bayer AG entered into a collaboration agreement in 1998 that gives Bayer a license to Exelixi's "FlyTag" Drosophila EST database and obligates Exelixis to develop new pest species EST databases for Bayer (JA0365-73);

- Human Genome Sciences has licensed its proprietary EST databases to companies and public scientific institutions. (JA0410-14.)

ESTs also are used for the study of gene expression in the burgeoning field of microarray analysis. (JA0078:24-26.)[10] According to recent reports, the global microarray market is poised to grow to nearly $1 billion in annual revenues by 2010. *See, e.g.*, Frost & Sullivan, *Strategic Analysis of World DNA Microarray Markets*, Report A776, Mar. 30, 2004.

---

[10] Microarrays display ordered sets of data points that correspond to known DNA molecules. Scientists can use microarrays to detect thousands of genes in a small sample simultaneously, and to analyze the expression of those genes. (JA0077:25-JA0078:14; JA0078:24-26.)

C. The Invention of Claim 1 of the '643 Application

The invention of claim 1 of the '643 Application is directed to certain purified nucleic acid molecule sequences that encode proteins, and fragments of proteins, produced in maize plants. (JA0028:5-7.) Claim 1 provides as follows:

> A substantially purified nucleic acid molecule that encodes a maize protein or fragment thereof comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 though SEQ ID NO: 5.

(JA0114:2-4; JA0169.)[11] The Applicants derived the five claimed nucleic acid molecules ("the claimed ESTs")[12] from a random sampling of cDNA clones generated from maize pooled leaf tissue at the time of anthesis. (JA0057:21-24; JA0106:18-JA0107:6.)[13]

1. The Claimed ESTs Can Be Used as Research Probes.

At the time the Applicants filed the '643 Application, the claimed ESTs were known to correspond to genes expressed in maize leaf tissue at the time of anthesis; however, the specific functions of the expressed genes corresponding to

---

[11] As used in claim 1, the term "substantially purified" refers to "a molecule separated from substantially all other molecules normally associated with it in its native state." (JA0042:19-25.)

[12] The invention of claim 1 covers ESTs, as well as other purified nucleic acid molecules falling within the scope of the claim. The Applicants' reference herein to "the claimed ESTs" is solely intended as a convenient form of shorthand and in no way is intended to serve as a disclaimer of claim scope.

[13] "Anthesis" refers to "the period during which a flower is fully open and functional." AMER. HERITAGE DICTIONARY OF THE ENGLISH LANGUAGE 4th Ed. (2000).

the claimed ESTs were unknown. (JA0057:21-24.) Nevertheless, even in the absence of that functional information, like any EST, each of the claimed ESTs could be used as research probes to screen a cDNA library for the specific gene sequence to which the EST uniquely corresponds. (JA0077:18-24.) The successful hybridization between one of the claimed ESTs and its corresponding gene sequence would confirm that the gene was being expressed in certain tissues, at certain times, by certain organisms. (JA0074:11-22; JA0077:25-JA0078:5; JA0273-78.)

The information derived from EST probes is useful to geneticists in a number of key respects, even where, as here, the specific function of the corresponding gene is unknown.[14] For example, as detailed below, each of the claimed ESTs can be used as research tools to: (1) serve as molecular markers on a genetic or physical map; (2) measure the level of mRNA in a sample; (3) serve as a source for primers; (4) identify the presence or absence of a polymorphism; (5) isolate promoters; (6) control the expression levels of protein; and (7) locate genetic molecules of other plants and organisms. These uses provide the scientific

---

[14] Indeed, it is standard practice in the field of genetics to use EST sequences to screen cDNA libraries for expressed genes without undertaking the time consuming and economically burdensome task of sequencing and characterizing the function of each and every located target gene. (JA0077:25-JA0078:5; JA0273-78.)

community with a unique molecular tool for the targeting and isolation of novel genes for plant protection and development.

2. Use of the Claimed ESTs as Molecular Markers

The claimed ESTs can be used as molecular markers on physical or genetic maps without knowing ***anything*** about the function of the corresponding gene. (JA0041:11-14; JA0050:1-JA0052:23; JA0074:1-JA0076:8; JA0077:18-JA0080:18.) Geneticists searching the maize genome utilize maps to guide them along through the more than two billion base pairs found in the DNA of each plant cell. However, for a map to make navigational sense, it must include reliable landmarks or markers that help determine the order of genes and distances between sequences. (JA0069:19-JA0073:27.)

The claimed ESTs can do just that. When one of the claimed ESTs is introduced into a genetic sample and hybridizes with its target complementary DNA sequence, the specific location where the probe hybridizes can serve as a molecular marker on a physical or genetic map. (JA0069:19-JA0070:12; JA0074:1-2; JA0074:23-JA0076:8.) This landmark information is useful to scientists – even in the absence of information about the function of the gene corresponding to the EST. For instance, when considered in connection with other markers, the presence or absence of a molecular marker corresponding to one of the claimed ESTs can be used to help determine the pedigree of a plant and

estimate the likely traits of a plant. (JA0069:19-JA0070:12.) Molecular markers also are useful in other applications as well, including linkage analysis, marker-assisted breeding, transgenic crop production, crop monitoring, and diagnostics. (JA0069:19-JA0076:8.)

3. Use of the Claimed ESTs to Measure mRNA Levels

The claimed ESTs also can be used to confirm the presence and quantitative measurement of an mRNA molecule within a particular tissue or cell sample. To do so, the claimed ESTs are used as probes to screen a sample of genetic material. (JA0074:11-JA0076:12.) "The detection of double-stranded probe-mRNA hybrid molecules is indicative of the presence of the mRNA." (JA0074:16-17.) "[T]he amount of such hybrid formed is proportional to the amount of mRNA" in the sample. (JA0074:17-18.) Thus, the claimed ESTs "may be used to ascertain [whether mRNA is present, and if so], the level and extent of the mRNA production in a plant's cells or tissues." (JA0074:18-19.) This information can be used to identify the type or source of a particular tissue, or to help evaluate how a plant's cells or tissues respond in a particular setting, such as when the plant is infected with disease or subjected to adverse growing conditions. (JA0074:11-JA0076:8.)

4. Use of the Claimed ESTs as a Source of Primers

The complex process necessary to sequence a gene or gene fragment requires many copies of the target DNA molecule. Without knowledge of the underlying gene function, the claimed ESTs can be used to help implement a "polymerase chain reaction" (PCR) – a well-known method that can generate billions of copies of a target DNA molecule within a matter of hours. (JA0060:7-13; JA0064:13-20; JA0277.)

PCR involves heating a DNA sample to separate the double-stranded target DNA molecule into two single strands. When the mixture cools, a primer in the sample will anneal (i.e., bind) to its complementary sequence on the first strand, and a second primer will anneal to its complementary sequence on the second strand.[15] DNA polymerase (an enzyme that catalyzes the synthesis of nucleic acids) then is used in conjunction with the annealed primers to synthesize two new DNA strands that are complementary to the original two strands. The two newly created strands anneal to the two original strands of the target DNA molecule, resulting in two complete target DNA molecules. When the two target DNA

---

[15] A primer is a short, single-stranded DNA molecule that is complementary to the sequence found at one end of the target DNA strand. The short sequence typically is unique to the target DNA molecule; therefore, when introduced into a sample, the primer will anneal ***only*** to the target DNA molecule. In nature, primers are formed from the free nucleotides residing in a cell by an enzyme called DNA primase. Primers also can be synthesized in a lab environment. (JA0052:24-JA0053:5; JA0277.)

molecules are subjected to another cycle of heating, the strands of both DNA molecules separate, resulting in four strands that each becomes a template for DNA replication using the primer method discussed above. This heating and cooling process is continued as necessary, with each cycle doubling the amount of target DNA. (JA0060:7-13; JA0064:8-JA0065:2; JA0277.)

The claimed ESTs can be used to save valuable time and effort needed to generate primers for use in the PCR process. As noted above, ESTs – like primers – typically represent the coding sequence found at the end of one of the strands of a specific DNA molecule. Therefore, the claimed ESTs can serve as a readily available template to design primers specific to a given gene, thereby allowing scientists to generate large sample populations of the corresponding gene sequences in a rapid and cost-efficient manner. (JA0060:7-13; JA0277.)

5. Use of the Claimed ESTs to Detect the Presence or Absence of a Polymorphism

The claimed ESTs also can be utilized as probes to identify the presence or absence of a polymorphism in a population of maize plants. (JA0062:20-JA0069:6.)[16] The presence of a polymorphism indicates that the respective genetic codes of the two or more samples being compared differ in some respect.

---

[16] "The genomes of animals and plants naturally undergo spontaneous mutation in the course of their continuing evolution." (JA0063:8-9.) A "polymorphism" is "a variation or difference in the sequence of the gene or its flanking regions that arises in some of the members of a species." (JA0063:9-11.)

Therefore, use of the claimed ESTs to confirm the existence of a polymorphism provides information that can be used, for instance, to enable a plant breeder to determine the distribution of parental genetic material passed to a newly bred plant. Polymorphic information also is useful to relate a particular genetic deviation to a particular observable trait for purposes of tracking the trait or predicting the likelihood of the trait being present or absent in progeny plants. (JA0062:1-23; JA0066:9-11; JA0069:7-JA0074:2.)[17]

The claimed ESTs also can be used to confirm the absence of a polymorphism. (JA0062:1-23.) This finding typically demonstrates that the two or more populations being compared share a common genetic heritage, and is useful, for example, to confirm that a parental plant passed a gene corresponding to one of the claimed ESTs to a progeny plant. (JA0069:7-JA0070:2.) Confirming the absence of a polymorphism also is useful in constructing genomic maps and assessing relationships between various traits and polymorphic markers. (JA0069:7-23.)

---

[17] Once a polymorphism is discovered, the claimed ESTs can serve as markers that are genetically or physically linked to the polymorphic area. (JA0063:3-7; JA0069:7-23; JA0074:1-2.) This information is "useful, through linkage analysis, to define the genetic distances or physical distances between polymorphic traits," "to characterize and isolate genes corresponding to desirable traits," and "to predict the phenotype [or traits] of the plant." (JA0066:9-11; JA0069:7-10.)

6. Use of the Claimed ESTs to Isolate Promoters

A "promoter" is a specific region of a gene that regulates the expression of protein. (JA0060:14-JA0061:13.) As described in the '643 Application, the claimed ESTs can be used to isolate promoters in specific tissue, including the promoter that regulates the expression of protein in maize leaf tissue at the time of anthesis by the genes that correspond to the claimed ESTs. (JA0060:14-JA0061:26.)

Promoter isolation may be accomplished by using techniques such as chromosome walking – a process that utilizes a known fragment of DNA (in this case, the claimed ESTs) to isolate adjacent fragments of DNA. (JA0060:20-25.) When used is this manner, the claimed ESTs are introduced as probes into a genomic library to screen for all clones that hybridize with the probe. The located clone that extends furthest into the surrounding DNA then is used as a probe on more distal regions of the DNA. The process then is repeated to "walk" down the target region of the chromosome. Chromosome walks using the claimed ESTs can help, for example, to sequence a DNA molecule or create a physical map of the maize genome. (JA0060:17-JA0061:8.)

7. Use of the Claimed ESTs to Control the Expression Levels of Protein

As disclosed in the specification of the '643 Application, the claimed ESTs can be used to control the expression levels of protein by a gene. (JA0061:22-26;

JA0099:9-JA0101:16.) For example, the claimed ESTs can be introduced into plant tissue or cells to inhibit or reduce the expression of a protein by the corresponding genes. (JA0099:14-JA0100:16.) Conversely, the claimed ESTs also can be introduced into a sample to induce greater expression levels of a protein by those genes. (JA0061:22-26; JA0099:9-14.) The ability to control protein expression levels using the claimed ESTs allows scientists to monitor how plant cells behave when the protein level is eliminated, reduced, or exaggerated. The resulting protein expression patterns aid scientists in understanding the function of the expressed gene and how to affect the pathways that regulate disease and other plant traits. (JA0057:21-JA0058:6; JA0061:25-26.)

8. Use of the Claimed ESTs to Locate Genetic Molecules of Other Plants and Organisms

Many different organisms contain genes that express proteins that are the same as, or substantially similar to, those expressed by genes found in maize plants. The claimed ESTs can be used to probe genetic libraries for such gene sequences of interest found in other organisms. (JA0058:23-JA0060:13; JA0060:17-JA0061:8.) In particular, the '643 Application expressly discloses that the claimed ESTs may be used to isolate nucleic acid molecules found in organisms such as alfalfa, *Arabidopsis*, barley, cotton, oat, oilseed, rape, rice, and many others. (JA0059:11-23.)

If one of the claimed ESTs hybridizes with a gene sequence found in another organism, and the gene has a known function with respect to the other organism, that finding may serve as a shortcut to help sequence the full-length gene and determine how the gene functions in maize plants. (JA0059:11-18.) Similarly, the knowledge derived from hybridization between one of the claimed ESTs and a gene sequence found in another organism is important even in the absence of knowledge about the gene's function in the other organism. For example, such a correlation suggests that the organisms under study may share a common genetic heritage. (JA0059:11-JA0060:13.) For this reason, the mere knowledge that a gene corresponding to one of the claimed ESTs exists in a different organism alone provides geneticists with valuable information.

D. The Board's Decision on Appeal

On appeal, the Board upheld the Examiner's rejection of claim 1 based on lack of utility and enablement grounds, but reversed the Examiner's rejection for lack of written description. These grounds are detailed below.[18]

---

[18] Before rejecting claim 1, the Board construed it to cover "a nucleic acid molecule, separated from substantially all other molecules normally associated with it in its native state, selected from the group consisting of the nucleic acid molecule defined by the 429 nucleotide sequence set forth in SEQ ID NO: 1, the 413 nucleotide sequence set forth in SEQ ID NO: 2, the 365 nucleotide sequence set forth in SEQ ID NO: 3, the 414 nucleotide sequence set forth in SEQ ID NO: 4, and the 333 nucleotide sequence set forth in SEQ ID NO: 5, with or without any preceding or trailing nucleotides, or other molecules." (JA0005.)

1. The Board's Utility Rejection

The Board conceded on appeal that the claimed ESTs can be used as probes in a variety of different scientific applications. (JA0014; JA0016; JA0021; JA0023-24.) Nevertheless, the Board concluded that those uses are insufficient, either alone or in combination, to satisfy the utility requirement of 35 U.S.C. § 101.

In assessing the utility of the claimed ESTs, the Board announced and then applied a new heightened standard of utility under Section 101 – a standard that hinges upon some undefined "spectrum" of knowledge concerning the function of a gene corresponding to an EST:

> Somewhere between having no knowledge (the present circumstances) and having complete knowledge of the gene and its role in the plant's development and/or phenotype lies the line between "utility" and "substantial utility."

(JA0015.) The Board expressly refused to provide any guidance with respect to the application of its newly crafted test, however, other than to conclude that the test was not met here: "We need not draw the line or further define it in this case because the facts in this case represent the lowest end of the spectrum, i.e., an insubstantial use." (JA0015-16.)

The Board also made several additional findings in an effort to support its lack of utility determination.

- First, the Board attempted to equate the claimed ESTs with the chemical compositions of ***no known function*** at issue in cases such as *Brenner v. Manson*. (JA0019; JA0022.)
- Second, the Board determined that uses of the claimed ESTs resulting in just a ***single data point*** (e.g., molecular markers or to measure mRNA levels) lack patentable utility because "[p]roviding a single data point among thousands or millions, even if the thousands or millions of data points collectively are useful, does not meet this standard [for utility]." (JA0022-24.)
- Third, the Board concluded that legal utility cannot be founded upon uses of the claimed ESTs that require ***additional testing*** to interpret the information derived from such uses (e.g., locating polymorphisms and promoters, or to measure the level of mRNA). (JA0014; JA0021-22.)
- Fourth, the Board found that the specification failed to disclose a ***reasonable expectation*** that the claimed ESTs could be used to isolate promoters. (JA0017.)
- Fifth, the Board characterized arguments directed to the ***multi-million dollar industry*** in the United States premised upon the usefulness of ESTs as irrelevant, based on its finding that the industry is founded

upon the value of EST databases, clone sets, and microarrays, and not upon single ESTs standing alone. (JA0024.)

2. The Board's Enablement Rejection

The Board upheld the Examiner's rejection of claim 1 for lack of enablement under 35 U.S.C. § 112. However, the Board's enablement rejection hinged entirely upon its utility rejection: "This rejection is simply a corollary of the finding of lack of utility." (JA0024-25.)

3. The Board's Reversal of the Examiner's Written Description Rejection

During prosecution of the '643 Application, the Examiner rejected claim 1 for lack of written description under 35 U.S.C. § 112, finding that the Applicants' use of "comprising" language resulted in coverage of nucleic acid molecules not described in the specification. The Board reversed the Examiner's ruling, finding "that the claimed [ESTs] may have other molecules attached ... does not diminish [Applicants'] adequate written description of the [claimed ESTs]." (JA0025-26.) Therefore, the Board found that the '643 Application contained a written description adequate to cover the claimed ESTs, as well as "inter alia, genes, full opening reading frames, fusion constructs, and cDNA." (JA0005; JA0025-26.)

IV. SUMMARY OF ARGUMENT

As the U.S. Supreme Court announced nearly four decades ago in *Brenner v. Manson*, the "basic quid pro quo contemplated by the Constitution and the

Congress for granting a monopoly is the benefit derived by the public from an invention with substantial utility" – that is, "where specific benefit exists in currently available form." *Brenner v. Manson*, 383 U.S. 519, 534-35 (1966). The Applicants have satisfied their end of the bargain here. The record contains more than ample evidence demonstrating that, as a matter of scientific truth, ***all*** ESTs – including each of the claimed ESTs – can be used as research tools to provide the public with a host of specific, substantial, and commercially valuable benefits – ***regardless*** of the level of knowledge of the corresponding gene function. This is all that the minimal utility requirement of 35 U.S.C. § 101 demands.

In upholding the Examiner's utility rejection, the Board failed to apply the proper standard of utility to the facts of this case. Instead, the Board fashioned new law by applying a heightened "spectrum" of knowledge utility test under Section 101 to reject the patentability of claim 1. In so doing, the Board inappropriately expanded the holding of *Brenner* and its progeny by effectively declaring that ***no*** EST can satisfy the utility requirement of Section 101 in the absence of some undefined level of knowledge concerning the function of the corresponding gene. This was error. Neither *Brenner* nor the U.S. patent laws support the stringent and vaguely defined utility analysis created and applied by the Board to reject claim 1.

At bottom, the Board erroneously concluded that claim 1 of the '643 Application fails to satisfy the minimal standard for legal utility established by 35 U.S.C. § 101. For these same reasons, the Board also erred in concluding that claim 1 does not meet the enablement requirement of 35 U.S.C. § 112; a finding grounded entirely upon the Board's lack of utility determination. Accordingly, the Court should reverse the Board's unfounded rejection of claim 1.

V. ARGUMENT

A. Standard of Review

The Board's lack of utility determination under 35 U.S.C. § 101 is a factual finding that the Court reviews for "substantial evidence." *See In re Gartside*, 203 F.3d 1305, 1315 (Fed. Cir. 2000) ("[W]e ... review Board factfinding for 'substantial evidence.'"); *In re Cortright*, 165 F.3d 1353, 1356 (Fed. Cir. 1999) ("Utility is a factual issue...."). This standard of review "involves examination of the record as a whole, taking into account evidence that both justifies and detracts from an agency's decision." *In re Gartside*, 203 F.3d at 1315.

The Board's enablement rejection under 35 U.S.C. § 112 is a question of law that this Court reviews *de novo*. *See In re Vaeck*, 947 F.2d 488, 495 (Fed. Cir. 1991) ("Enablement ... is a question of law which we independently review"). The Court reviews the factual findings underlying the Board's enablement rejection for "substantial evidence." *See In re Gartside*, 203 F.3d at 1315.

The Board's statutory interpretation of 35 U.S.C. §§ 101 and 112 is a legal issue that the Court reviews *de novo*. *See In re Carlson*, 983 F.2d 1032, 1035 (Fed. Cir. 1992) ("Interpretation of statutory terms is a question of law which this court reviews de novo.") However, in construing Sections 101 and 112, the Court should be mindful "not [to] read into the patent laws limitations and conditions which the legislature has not expressed." *Diamond v. Chakrabarty*, 447 U.S. 303, 308 (1980).

B. The Utility Requirement Established by Congress Has a Minimal Threshold.

The concept of utility embodies "a fundamental requirement of American patent law" that finds its roots and purpose in the United States Constitution: That "[t]he Congress shall have power ... to promote the progress of science and the ***useful*** arts." U.S. CONST., ART. I § 8 (emphasis supplied); *see Stiftung v. Renishaw PLC*, 945 F.2d 1173, 1180 (Fed. Cir. 1991) ("The utility requirement has its origin in article I, section 8 of the Constitution, which indicates that the purpose of empowering Congress to authorize the granting of patents is 'to promote progress of ... *useful* arts.'").

1. The Utility Requirement of 35 U.S.C. § 101

Over the last two centuries, Congress has enacted a regime of patent laws to "promote this progress by offering inventors exclusive rights for a limited period as an incentive for their inventiveness and research efforts." *Chakrabarty*, 447

U.S. at 307. Today, the constitutional requirement of "useful" inventions is codified by 35 U.S.C. § 101, which provides in relevant part that:

> whoever invents or discovers any new and ***useful*** process, machine, manufacture, or composition of matter, or any ***useful*** improvement thereof, may obtain a patent therefor . . . .

35 U.S.C. § 101 (emphasis supplied).

The legislative history unambiguously indicates that Congress intended for the "extremely broad" language of Section 101 to "be given wide scope" and "a broad construction" so as to cover "anything under the sun that is made by man." *See J.E.M. AG Supply, Inc. v. Pioneer Hi-Bred Int'l, Inc.*, 534 U.S. 124, 130 (2001); *Chakrabarty*, 447 U.S. at 308-09; S. REP. NO. 1979, 82d Cong., 2d Sess., 5 (1952); H.R. REP. NO. 1923, 82d Cong., 2d Sess., 6 (1952). The goal of this legislative scheme was to foster "a positive effect on society through the introduction of new products and processes of manufacture into the economy, and the emanations by way of increased employment and better lives for our citizens." *Kewanee Oil Co. v. Bicron Corp.*, 416 U.S. 470, 480 (1974).

Given the expansive breadth of this statutory language and congressional intent, courts construing Section 101 (and predecessor versions of that statute) historically have ascribed a minimal standard to the requirement that an invention be "useful." For example, in a well-known decision now nearly two centuries old, Justice Story announced that:

> All the law requires is that the invention should not be frivolous, or injurious to the well-being, good policy, or good morals of society. The word *useful* therefore is incorporated into the act in contradistinction to mischievous or immoral.

Lowell v. Lewis, 15 F. Cas. 1018, Case No. 8568 (C.C. Mass. 1817); *see also Evans v. Eaton*, 16 U.S. 454, 518 (1818) ("By useful invention, in the patent act, is meant an invention which may be applied to a beneficial use in society, in contradistinction to an invention injurious to the morals, health, or good under of society, or frivolous and insignificant."). Even today, courts – including this Court – continue to cite Justice Story's minimalist view of utility with approval. *See, e.g., Juicy Whip, Inc. v. Orange Bang, Inc.*, 185 F.3d 1364, 1366-67 (Fed. Cir. 1999) (noting that "[c]ourts have continued to recite Justice Story's formulation," but explaining that the prohibition on patenting inventions "principally designed to serve immoral or illegal purposes has not been applied broadly in recent years").

2. The Requirement of "Substantial" or "Practical" Utility Established by *Brenner v. Manson*.

Nearly four decades ago, a divided Supreme Court addressed the standard of utility applicable to a chemical composition with no known utility "other than as a possible object of scientific inquiry." *Brenner*, 383 U.S. at 529. After acknowledging the sometimes-difficult task of assessing the utility of inventions directed to chemical compositions, the *Brenner* court determined that to be

"useful" and patentable, inventions must provide the public with at least one identifiable benefit that is "substantial:"

> The basic quid pro quo contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with ***substantial utility***.

Id. at 534-35 (emphasis supplied). The Supreme Court further clarified that "substantial utility" does not exist "[u]nless and until ... specific benefit exists in currently available form." *Id.* Otherwise, "there is insufficient justification for permitting an applicant to engross what may prove to be a broad field." *Id.* "[A] patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion." *Id.* at 536.

In the several decades since *Brenner*, this Court and its predecessor have implemented a two-prong analytical framework to assess whether an invention provides "substantial utility." Under that standard, the claimed invention first must provide at least one specific, "identifiable benefit" – i.e., one that is not vague or unknown. *See Juicy Whip*, 185 F.3d at 1366; *In re Brana*, 51 F.3d 1560, 1565 (Fed. Cir. 1995); *Application of Kirk*, 376 F.2d 936, 945 (C.C.P.A. 1967).[19]

---

[19] The utility requirement of Section 101 demands only ***one*** actual, identifiable benefit. *See Stiftung*, 945 F.2d at 1180 ("[W]hen a properly claimed invention meets at least one stated objective, utility under § 101 is clearly shown;" to satisfy the utility requirement does not "mean that a patented device must accomplish *all* objectives stated in the specification.").

This Page Blank (uspto)

Second, the benefit must also be "substantial" or "practical" – i.e., one that provides a measurable benefit in the "real world." *See Fujikawa v. Wattanasin*, 93 F.3d 1559, 1563 (Fed. Cir. 1996); *Nelson v. Bowler*, 626 F.2d 853, 856 (C.C.P.A. 1980).[20]

3. Even After *Brenner*, "The Threshold Of Utility Is Not High."

The Board's Decision appears to suggest that *Brenner* marked a radical departure from Justice Story's minimalist view of utility. It did not. The decisions of this Court and its predecessor repeatedly have confirmed that the threshold to demonstrate "substantial utility" under Section 101 remains strikingly minimal, even today:

- Lack of utility is shown "when there is ***a complete absence of data*** supporting the statements which set forth the desired results of the claimed invention." *In re Cortright*, 165 F.3d at 1356 (emphasis supplied).
- "The threshold of utility ***is not high***: An invention is 'useful' under section 101 if it is capable of providing ***some identifiable benefit***." *Juicy Whip*, 185 F.3d 1366 (citing *Brenner v. Manson* as direct support for holding) (emphasis supplied).
- "To violate § 101 the claimed device must be ***totally incapable of achieving a useful result***." *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555, 1571 (Fed. Cir. 1992) (emphasis supplied).

---

[20] The Court has long-treated the terms "substantial utility" and "practical utility" interchangeably. *See, e.g., Fujikawa*, 93 F.3d at 1563-64; *Cross v. Iizuka*, 753 F.2d 1040, 1047 n.13 (Fed. Cir. 1985).

- "[A] reasonable jury could not have found the ***'total incapacity'*** that is required to prevail on a lack of utility defense under § 101." *Tol-O-Matic, Inc. v. Proma Produkt-Und Marketing Gesellschaft m.b.H*, 945 F.2d 1546, 1553 (Fed. Cir. 1991) (emphasis supplied).

- "***Some degree of utility*** is sufficient for patentability." *Envirotech Corp. v. Al George, Inc.*, 730 F.2d 753, 762 (Fed. Cir. 1984) ("[T]he defense of non-utility cannot be sustained without ***proof of total incapacity***.") (emphasis supplied).

- "Practical utility is a shorthand way of attributing 'real-world' value to the claimed subject matter. In other words, one skilled in the art can use a claimed discovery in a manner which provides ***some immediate benefit*** to the public." *Nelson*, 626 F.2d at 856 (emphasis supplied).

The few post-*Brenner* cases actually finding lack of utility further confirm that the standard for utility remains minimal. *See, e.g., Newman v. Quigg*, 877 F.2d 1575, 1577, 1581-82 (Fed. Cir. 1989) (finding lack of utility where "perpetual motion machine" deemed "impossible" under the laws of thermodynamics); *Fregeau v. Mossinghoff*, 776 F.2d 1034, 1039 (Fed. Cir. 1985) (finding no utility for method to enhance the flavor of beverages using a magnetic field); *Application of Houghton*, 433 F.2d 820, 820-21 (C.C.P.A. 1970) (finding no utility for "highly unusual" flapping flying machine based on bird and insect flight); *Application of Eltgroth*, 419 F.2d 918, 920-21 (C.C.P.A. 1970) (no utility for "speculative" method of controlling the aging process); *Kirk*, 376 F.2d at 942-43 (utility lacking for steroid compound where disclosed "possible use so general as to be meaningless").

C. The Board Improperly Applied a Heightened and Ill-Defined Standard of Utility to Reject the Patentability of the Claimed ESTs.

Rather than apply the minimal standard of utility established by Section 101 and repeatedly applied by the courts, the Board instead applied a new heightened utility standard of its own creation that conditions the patentability of ESTs upon some undefined "spectrum" of knowledge concerning the corresponding gene function. (JA0015-16.) The Board's unilateral establishment of a utility requirement for ESTs that is far more demanding than the utility standard applicable to other chemical compounds and inventions was error.[21]

It is well established that Congress alone has been entrusted with the power to define the level of utility necessary to effectuate the constitutional requirement that patentable inventions be "useful;" a power that Congress has exercised through its enactment of 35 U.S.C. § 101. *See, e.g.*, U.S. CONST., ART. I § 8; *Chakrabarty*, 447 U.S. at 307; *Juicy Whip*, 185 F.3d at 1368. However, nothing in the plain language of Section 101 or its legislative history in any way supports a claim that Congress expressly or impliedly intended to subject ESTs to a utility

---

[21] Acceptance of the Board's heightened utility standard would lead to unjustifiable results. By way of example, under the test articulated by the Board, a drink dispenser whose sole benefit is that it "look[s] like another product" will have patentable utility, *see Juicy Whip*, 185 F.3d at 1367, while certain ESTs will be rejected for lack of utility even though they can be used to help scientists further understand the complex genome of the scientifically and commercially important maize plant.

standard that is more stringent than the standard applicable to other inventions. In fact, the decisions of this Court make abundantly clear that just the opposite is true. *See, e.g., Juicy Whip*, 185 F.3d at 1368 (applying the same "substantial utility" standard to an imitation drink dispenser that the Supreme Court applied to the chemical composition at issue in *Brenner*).

The Board simply does not have the authority or the expertise to usurp the role of Congress by rewriting the statutorily mandated standard for utility applicable under 35 U.S.C. § 101. *See Chakrabarty*, 447 U.S. at 308 (noting that courts should be mindful "not [to] read into the patent laws limitations and conditions which the legislature has not expressed"). Accordingly, rejection of the Board's newly manufactured "spectrum" of knowledge utility test is warranted.[22]

D. The Claimed ESTs Meet the Utility Requirement of 35 U.S.C. § 101.

Under a proper application of the law, there is no question that the claimed ESTs satisfy the threshold for utility established by Section 101 – which "is not high." *See Juicy Whip*, 185 F.3d at 1366. The record on appeal confirms that the claimed ESTs are important research tools that can be put to a variety of specific,

---

[22] In recent months, the Board has applied its heightened "spectrum" of utility standard to reject a number of other patent applications directed to ESTs on utility grounds. *See* Statement of Related Cases, *supra*, at viii. In order to avoid duplicative appeals on this same issue and to provide needed certainty to the patenting process, Applicants urge the Court to not only reject the new legal standard devised by the Board, but also to clarify the proper standard to be applied in determining the utility of ESTs in future cases.

This Page Blank (uspto)

substantial, and commercially beneficial uses beyond mere use-testing. These disclosed uses require reversal of the Board's utility rejection.

1. The Claimed Invention is Entitled to a Presumption of Utility.

The disclosure in the '643 Application of a myriad of uses for the claimed ESTs gives rise to a presumption that the utility requirement of 35 U.S.C. § 101 has been met. *See In re Brana*, 51 F.3d at 1566 ("[T]he PTO has the initial burden of challenging a presumptively correct assertion of utility in the disclosure."). To overcome this presumption of utility, the Board must "provide[] evidence showing that one of ordinary skill in the art would reasonably doubt the asserted utility." *Id.* Only then "does the burden shift to the applicant to provide rebuttal evidence sufficient to convince such person of the invention's asserted utility." *Id.* As demonstrated in detail below, the Board has not met – and cannot meet – this stringent standard. Therefore, the claimed ESTs are entitled to a presumption (and finding) of utility.[23]

2. The Benefits Derived from Use of the Claimed ESTs Are Specific.

Even in the absence of a presumption of utility, the record confirms that ***all*** ESTs – including the claimed ESTs at issue here – meet the first prong of this

---

[23] In the proceedings below, the Board failed to apply (or even mention) the presumption of utility applicable to the Applicants' invention. This was error. *See In re Brana*, 51 F.3d at 1566.

Court's utility analysis by providing the public with a number of specific, identifiable benefits that are not vague or unknown.

More specifically, this is not a case about a chemical composition like that at stake in *Brenner,* which had ***no*** known identifiable use other than as a subject of further scientific research. Rather, the record in this case makes clear that ***all*** ESTs correspond to a specific gene with a ***knowable*** function. Furthermore, ***all*** ESTs can be used as important research tools in connection with a host of scientific applications, including to: (1) serve as molecular markers on a genetic or physical map; (2) measure the level of mRNA in a sample; (3) serve as a source for primers; (4) identify the presence or absence of a polymorphism; (5) isolate promoters; (6) control the expression levels of protein; and (7) locate genetic molecules of other plants and organisms.

Each of these disclosed uses is "identifiable" and specific to the claimed ESTs. In fact, because each of the claimed ESTs uniquely corresponds to a specific gene segment found in maize plants, no other EST can be utilized for exactly the same purposes. *See* Manual of Patenting Examining Procedure (MPEP) § 2107.01 at 2100-32 (defining "specific utility" as "[a] utility that is *specific* to the subject matter claimed," and contrasting it with "a *general* utility that would be applicable to the broad class of the invention.").)

Accordingly, the claimed ESTs meet the specificity prong of the Court's substantial utility analysis. *See Juicy Whip*, 185 F.3d at 1366 ("An invention is 'useful' under section 101 if it is capable of providing ***some identifiable benefit***."); *Nelson*, 626 F.2d at 856 (finding that utility exists where "one skilled in the art can use a claimed discovery in a manner which provides ***some immediate benefit*** to the public") (emphasis supplied).

3. The Benefits Derived From Use of the Claimed ESTs Are Substantial.

The record further confirms that, as a matter of scientific reality, ***all*** ESTs satisfy the "substantial benefit" prong of this Court's utility analysis – even with respect to ESTs, such as those claimed here, that correspond to genes of unknown function. There is no dispute that, when used as research probes to screen genetic samples for particular genes and gene fragments, the claimed ESTs provide the public with a number of measurable benefits, including to:

- Serve as molecular markers for genes of interest, thereby assisting scientists to navigate through complex physical and genetic maps detailing the billions of base pairs found in the maize genome (JA0041:11-14; JA0050:1-JA0052:23; JA0069:19-JA0070:12; JA0074:1-JA0076:8; JA0077:18-JA0080:18; JA00273-79);

- Detect and monitor the quantitative levels and patterns of mRNA found in a particular cell or tissue sample, thus providing information pertinent to detecting expression changes in plant traits of interest (e.g., drought stress) (JA0074:11-JA0076:8; JA0076:9-JA0077:17);

- Serve as a source of a synthetic PCR primer to enable the rapid and inexpensive duplication of a specific target gene (JA0052:24-JA0053:5; JA0060:7-13; JA0064:8-JA0065:2; JA0277);

- Determine the presence or absence of polymorphic variations between two or more populations of genetic samples, which, among other things, provides scientists with important information concerning the nature of any shared genetic heritage between the samples (JA0062:1-JA0074:2);

- Isolate promoters (such as the promoter of the genes corresponding to the claimed ESTs) by, for example, initiating a chromosome walk (JA0060:14-JA0061:26);

- Control the expression levels by a gene to allow study of protein expression patterns and gene protein function (JA0057:21-JA0058:6; JA0061:22-26; JA0099:9-JA0101:16); and

- Isolate nucleic acid molecules found in other organisms to allow comparative studies of located genes and their functions between organisms. (JA0058:23-JA0061:8.)[24]

As detailed above, the record confirms that each of these uses furnishes the field of genetic science with ***substantial*** benefits that are capable of realization in the real world – regardless of the level of knowledge concerning the function of the underlying gene. Accordingly, the Board's failure to credit these many substantial uses constitutes reversible error. *See, e.g., Juicy Whip*, 185 F.3d at 1366 (holding that "substantial utility" standard is met by an invention that "is capable of

---

[24] As this Court has recognized, in the absence of specific evidence to the contrary, each of these disclosed uses must be presumed under the law to be substantial. *See In re Brana*, 51 F.3d at 1566. Here, there is no such contrary evidence.

providing ***some identifiable benefit***."); *Nelson*, 626 F.2d at 856 (holding that substantial utility merely requires that "one skilled in the art can use a claimed discovery in a manner which provides ***some immediate benefit*** to the public") (emphasis supplied).

As a practical matter, a decision to the contrary would mean that other inventions of unquestionable and critical value to the scientific community similarly lack substantial utility. For example, in a number of key respects, the claimed ESTs are directly analogous to research tools such as microscopes, telescopes, and screening assays – all of which can be utilized to study, locate, and generate scientific data about samples with currently unknown properties. It would make little sense to hold – as the Board effectively did below – that research tools such as these have substantial utility when used to analyze a sample of known function, but no substantial utility when used to analyze a sample of unknown function. In fact, research tools arguably have even greater value when used to

probe, examine, and understand the properties of a sample with an unknown function.[25]

The same holds true here with respect to the invention of the '643 Application. When used as a probe to screen a genetic sample, each of the claimed ESTs can be used like a microscope to locate, study, and derive information about particular genes or gene fragments. That the gene under examination has no known function does not change this result. Like a microscope, regardless of whether used to examine genes of ***unknown*** function now, or genes of ***known*** function at some later date, the claimed ESTs serve specific, substantial, and scientifically valuable purposes. They are patentable.

4. The Patentable Utility of the Claimed ESTs is Further Confirmed by Considerations of Commercial Success.

As the Supreme Court noted decades ago in *Brenner,* the test for utility is closely "related to the world of commerce rather than to the realm of philosophy." *See Brenner*, 383 U.S. at 536. For that reason, "[p]roof of ... utility is further supported when ... the inventions ... have on their merits been met with

---

[25] On appeal, the Board found that the claimed ESTs differ from a microscope "because the microscope provides information to the scientist which is automatically useful. For example, the microscope may be used for identification and differentiation between gram-positive and gram-negative bacteria." (JA0015.) Notably, however, the Board never suggested that a microscope would lack patentable utility when used to examine newly discovered forms of bacteria of unknown function. Tellingly, nor did the Board provide any evidence to support its incorrect factual conclusion that information gained from an EST is not automatically useful. As described above, it is.

commercial success." *Raytheon Co. v. Roper Corp.*, 724 F.2d 951, 959 (Fed. Cir. 1983).[26] This nexus between utility and commercial success exists because "[p]eople rarely, if ever, appropriate useless inventions." *Id.*; *see Application of Langer*, 503 F.2d at 1393.

Here, the utility of the claimed ESTs is not merely an abstract exercise in "the world of philosophy." Rather, a vast industry has developed in the commercial marketplace for ESTs – including for ESTs that, like those at issue here, code for genes of unknown function. Indeed, numerous companies have dedicated substantial funds to research and discover ESTs that correspond to genes of ***both*** known and unknown function, and large sophisticated companies collectively have paid hundreds of millions of dollars to obtain access to databases of ESTs. (JA0336-38; JA0341-46; JA0349-51; JA0355; JA0365-73; JA0377-78; JA0382; JA0396-403; JA0406-09; JA0410-14.)

It runs contrary to common sense to think that sophisticated corporations and knowledgeable scientists would dedicate hundreds of millions of dollars to an industry based upon useless items of commerce. Just as "[p]eople rarely, if ever, appropriate useless inventions," *Raytheon*, 724 F.2d at 949, people rarely, if ever,

---

[26] The reverse is not true: "development of a product to the extent that it is presently commercially salable in the market place is not required to establish 'usefulness' within the meaning of § 101." *Application of Langer*, 503 F.3d 1380, 1393 (C.C.P.A. 1974).

invest hundreds of millions of dollars in industries built upon useless inventions. The undeniable existence of such an industry here further confirms that the claimed ESTs meet the minimal utility requirement of 35 U.S.C. § 101. The Board erred by dismissing this relevant, probative evidence.

On appeal, the Board suggested that the success of the EST industry lacks probative value because the industry "is premised on ... the potential usefulness of EST databases, clone sets or microarrays" and "the claims on appeal are not directed to EST databases, clone sets, and/or microarrays." (JA0024.) Of course, the databases, clone sets, and microarrays would be useless without each of the individual ESTs. Indeed, the PTO itself has recognized as much in other cases by allowing patents to issue for inventions directed to a single component that plainly must be used with other components to have any meaningful commercial value – for example, a patent on a single LEGO block. (JA0427-31.)

Moreover, by general operation of the PTO's own rules (*see* MPEP § 803.04) and as a specific result of the restriction requirement issued in this case (JA0146-50), the Applicants were precluded from claiming anything beyond five ESTs. The Board cannot have it both ways. Having precluded the Applicants from claiming the more than 4,000 ESTs properly disclosed in the '643 Application, the Board should not be permitted to ground its lack of utility finding on the fact that the Applicants' claim is limited to only a handful of ESTs.

5. This Court's Utility Decisions Directed to the Analogous Field of Pharmacological Compounds Further Confirm the Utility of the Claimed ESTs.

This Court has yet to address the two specific issues raised on appeal here: (1) what standard of utility under 35 U.S.C. § 101 is applicable to ESTs; and (2) whether that standard is met here by the claimed ESTs, which correspond to genes of unknown function. Nevertheless, the Court has issued several instructive decisions in the analogous field of pharmaceutical compounds. In those cases, the Court has held that the utility requirement of Section 101 can be met by compounds claiming a pharmacological effect in humans, even in the absence of evidence showing an *in vivo* pharmacological activity. These decisions weigh in favor of a finding of utility here. *See Cross,* 753 F.2d at 1050*; Nelson*, 626 F.2d at 856; *In re Jolles*, 628 F.2d 1322, 1327 (C.C.P.A. 1980).

For example, in *Cross*, the Court held that *in vitro* testing – which "is but an intermediate link in a screening chain which may eventually lead to the use of the drug as a therapeutic agent in humans" – itself "may establish a practical utility for the compound in question." *Cross*, 753 F.2d at 1051. In so holding, the Court reasoned that "[s]uccessful *in vitro* testing will marshal resources and direct the expenditure of effort to further *in vivo* testing of the most potent compounds, thereby providing an immediate benefit to the public, analogous to the benefit provided by the showing of an *in vivo* utility." *Id.; see Nelson*, 626 F.2d at 856-57

(finding evidence of *in vivo* utility sufficient to establish utility); *In re Jolles*, 628 F.2d at 1327 (holding that animal testing satisfies the utility requirement).

The same rationale is applicable here. As disclosed in the specification of the '643 Application, the claimed ESTs can be utilized as tools to research the genetic profile of genes appearing in maize plants – even without knowledge of the gene functions corresponding to the ESTs. The information about a plant's genetic profile derived from use of the claimed ESTs, like the information about a compound's pharmacological profile in *Cross*, *Nelson*, and *In re Jolles*, provides an immediate and substantial benefit to the public. Indeed, use of the claimed ESTs to screen cDNA libraries for genes of interest found in the maize genome undoubtedly will help to "marshal resources and direct the expenditure of effort" with respect to further gene research (perhaps leading to improved plants) "thereby providing an immediate benefit to the public." *See Cross,* 753 F.2d at 1051.

6. The Other Stated Grounds Underlying the Board's Utility Rejection Lack Merit.

Beyond its improper announcement of a heightened "spectrum" of knowledge standard of utility applicable to ESTs, the Board further erred by making findings directed to the claimed ESTs that simply cannot withstand scrutiny. That is, despite conceding that the claimed ESTs can be used to screen

genetic libraries for genes and gene fragments of interest, as it must,[27] the Board advanced four arguments that purportedly support a finding that the claimed ESTs are incapable of providing the public with one or more specific benefits that are substantial. As discussed below, none of these arguments has merit.

First, the Board's attempt to equate the claimed ESTs with the chemical composition at issue in *Brenner* was misplaced. (JA0019; JA0022.) The *Brenner* chemical composition had ***no*** known utility other than as an object of further scientific research. In contrast, utility of the claimed ESTs does not rest upon some mere interest in conducting further research upon the ESTs themselves. Rather, utility exists here because the claimed ESTs – as well as all other ESTs – can be used as research tools to conduct further scientific research on ***other*** chemical compositions (i.e., genes, gene fragments, proteins, etc.).[28]

---

[27] The PTO's own rules acknowledge that nucleotide sequencing techniques ***are***, in fact, "useful in analyzing compounds." MPEP § 2107.01 at 2100-33 ("Many research tools such as ... nucleotide sequencing techniques have a clear, specific, and unquestionable utility (*e.g.*, they are useful in analyzing compounds).").)

[28] The *Brenner* majority held that a chemical composition "whose sole 'utility' consists of its potential role as an object of use-testing" does not have substantial utility. *Brenner*, 383 U.S. at 535. Writing for the dissent, Justice Harlan disagreed, noting that:

Moreover, unlike the chemical composition at issue in *Brenner* – which possibly had no actual discoverable utility at all – the claimed ESTs correspond to expressed genes that synthesize proteins a maize plant undoubtedly uses in leaves at the time of anthesis for some meaningful purpose.[29] The claimed ESTs can be used at a minimum to discover that unknown, but not unknowable, utility, as well as to conduct the variety of different research applications discussed above. These many uses collectively demonstrate that the claimed ESTs have significant value

---

> Chemistry is a highly interrelated field and a tangible benefit for society may be the outcome of a number of different discoveries, one discovery building upon the next. To encourage one chemist or research facility to invent and disseminate new processes and products may be vital to progress, although the product or process be without 'utility' as the Court defines the term, because that discovery permits someone else to take a further but perhaps less difficult step leading to a commercially useful item. In my view, our awareness in this age of the importance of achieving and publicizing basic research should lead this Court to resolve uncertainties in its favor and uphold the respondent's position in this case.

Id. at 539 (J. Harlan dissenting). Given the steep advances in the field of Chemistry over the last several decades, the Supreme Court might adopt Justice Harlan's dissenting view if asked to revisit the issue presented in *Brenner*. In any event, the Court need not reach that issue here inasmuch as the claimed ESTs have utility beyond as mere "objects of use-testing."

[29] "Leaves are the carbohydrate factories of crop plants, therefore, the ESTs of the present invention will find great use in the isolation of a variety of agronomically significant genes ...." (JA0057:24-26.)

beyond further "use-testing" of the EST sequences themselves. Thus, *Brenner* is wholly inapplicable on this point.[30]

Second, the Board erred in concluding that the benefits derived from certain uses of the claimed ESTs (e.g., as molecular map markers and to measure the level of mRNA in sample) are insubstantial because those uses result in the discovery of "a single data point among thousands or millions" of others. (JA0022-24.) Appellants are unaware of any case – and the Board cites none – which states that a research tool lacks patentable utility simply because it provides just one data point among many others, or because the results derived from that tool must be used in connection with other data to be completely meaningful.

Indeed, if accepted, the Board's rationale would have a profound impact on the patentability of numerous legitimate inventions that require the combination of multiple components or pieces of data to be completely useful. For example, if taken to its logical extreme, the Board's reasoning would preclude the patenting of inventions ranging from basic (e.g., a single LEGO block) to complex (e.g., gene

---

[30] Another case relied upon by the Board, *In re Ziegler*, 992 F.2d 1197 (Fed. Cir. 1993), is equally inapplicable. In that case, the Court found that a claimed polypropylene lacked utility even though it could be pressed into a film because Ziegler "did not assert any practical use for the polypropylene or its film, and Ziegler did not disclose any characteristics of the polypropylene or its film that demonstrated its utility." *Id.* at 1203. By contrast, the Applicants here have disclosed a number of practical uses for the claimed ESTs as research tools, none of which involves the use of ESTs as mere intermediates to generate a substance of no known value.

mapping, surface mapping, CAD modeling, or semiconductor fabrication systems) simply because those applications require the combination of many other components or data points to work and have value. For obvious reasons, this is not the law.

Third, the Board wrongly found that benefits derived from certain uses of the claimed ESTs are not "substantial" because additional testing might be required to make sense of the information derived from those uses. (JA0014; JA0021-22.) This rationale overlooks the cases of this Court holding that the need to conduct additional research to determine the ***significance*** of the particular results obtained from use of the claimed ESTs is no way fatal to the Applicants' invention. *See Cross,* 753 F.2d at 1051 (holding that information derived from *in vitro* test results were sufficient to meet the requirement of substantial utility, even though further research was required to prove ***any*** benefit for *in vivo* use); *Nelson*, 626 F.2d at 856 (finding utility without evidence of *in vivo* results); *In re Jolles*, 628 F.2d at 1327 (same).

In fact, as this Court has previously held, an "immediate benefit to the public" results from a composition that "marshal[s] resources and directs the expenditure of effort" for the very purpose of allowing ***additional testing***. *Cross,* 753 F.2d at 1051*; see Nelson*, 626 F.2d at 856 (finding evidence of *in vivo* utility sufficient to establish practical utility); *In re Jolles*, 628 F.2d at 1327 (holding that

animal testing satisfies the practical utility requirement). Similarly, information obtained from use of the claimed ESTs at issue here may permit researchers to identify genes and proteins of commercial importance more efficiently, leading to the development of maize strains that have improved characteristics.

Finally, the Board erroneously found that, although the claimed ESTs can be used to isolate promoters, the specification "fails to demonstrate that any of the [claimed] nucleic acid molecules ... would be useful in obtaining a successful result from such a search." (JA0017.) As an initial matter, the Board's reasoning reverses the presumption of utility that attaches to the Applicants' disclosure by forcing the Applicants to prove the correctness of their disclosed utility – without first offering evidence to the contrary. Nevertheless, even in the face of this improper burden shifting, the record confirms that the claimed ESTs can, in fact, be designed to isolate the promoter of the genes corresponding to the ESTs. (JA0060:14-JA0061:26.)

In sum, none of the stated grounds for the Board's utility rejection can withstand scrutiny. Accordingly, the Court should reverse the Board's unsupported utility rejection and allow claim 1 of the '643 Application to issue.

E. The Claimed ESTs Meet the Enablement Requirement of 35 U.S.C. § 112.

Although the Board cited lack of enablement under 35 U.S.C. § 112 as a separate ground for its rejection of claim 1, that rejection was entirely contingent

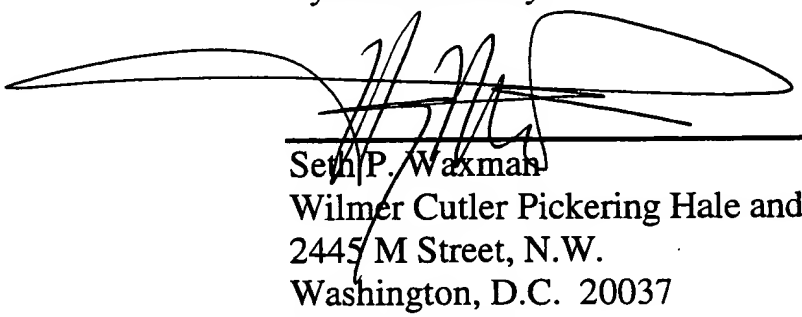
upon the Board's utility rejection. (JA0024-25.) Accordingly, because the Court should reverse the Board's lack of utility rejection, as detailed above, the Court should reverse the enablement rejection as well.

VI. CONCLUSION

For the foregoing reasons, the Court should reverse the Board's rejections for lack of utility and enablement under 35 U.S.C. §§ 101 and 112.

Appellants,
Dane K. Fisher and Raghunath V. Lalgudi

By their attorneys



Seth P. Waxman
Wilmer Cutler Pickering Hale and Dorr LLP
2445 M Street, N.W.
Washington, D.C. 20037
(202) 663-6800

William F. Lee
Richard W. O'Neill
Wilmer Cutler Pickering Hale and Dorr LLP
60 State Street
Boston, MA 02109
(617) 526-6497

William G. McElwain
Henry N. Wixon
Wilmer Cutler Pickering Hale and Dorr LLP
1455 Pennsylvania Ave., N.W.
Washington, D.C. 20004
(202) 942-8406

Dated: September 27, 2004

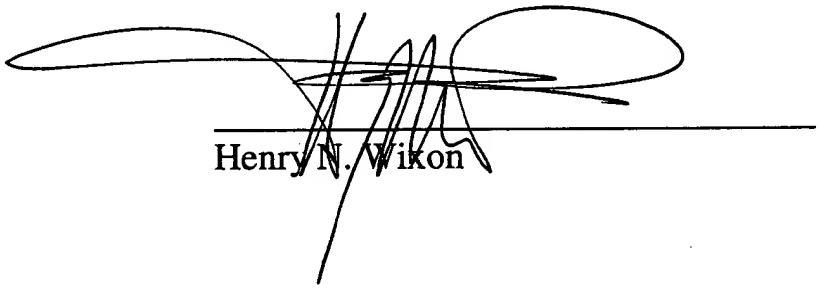
CERTIFICATE OF SERVICE

I, Henry N. Wixon, certify that on this 27th day of September, 2004, I caused two copies of the foregoing to be served by overnight mail upon the following:

John M. Whealan
Office of the Solicitor
P.O. Box 15667
Arlington, VA 22215

Stephen Walsh
Office of the Solicitor
P.O. Box 15667
Arlington, VA 22215

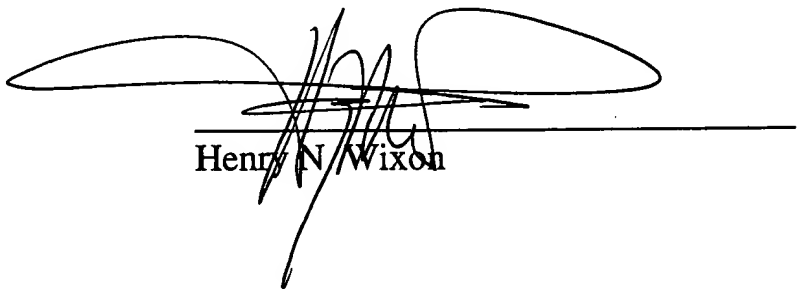
Thomas W. Krause
Office of the Solicitor
P.O. Box 15667
Arlington, VA 22215



Henry N. Wixon

CERTIFICATE OF COMPLIANCE WITH RULE 32(a)

I, Henry N. Wixon, certify that the foregoing brief complies with the type-volume limitation set forth in Fed. R. App. P. 32(a)(7)(B). Specifically, this brief contains 11,978 words, excluding the parts of the brief exempted by Fed. R. App. P. 32(a)(7)B)(iii), as determined by the word count feature of the word processing program used to create this brief. I further certify that the foregoing brief complies with the typeface requirements set forth in Fed. R. App. P. 32(a)(5) and the type style requirements of Fed. R. App. P. 32(a)(6). Specifically, this brief has been prepared using a proportionately spaced typeface using Microsoft 2000, in 14-point Times New Roman font.



Henry N. Wixon

This Page Blank (uspto)

ADDENDUM

Paper No. 17

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

Ex parte DANE K. FISHER, and RAGHUNATH V. LALGUDI

Appeal No. 2002-2046
Application No. 09/619,643

HEARD: March 16, 2004

Before WILLIAM F. SMITH, ADAMS, and GRIMES, Administrative Patent Judges.

ADAMS, Administrative Patent Judge.

DECISION ON APPEAL

This is a decision on the appeal under 35 U.S.C. § 134 from the examiner's final rejection of claim 1, the only claim pending in the application, reproduced below:

1. A substantially purified nucleic acid molecule that encodes a maize protein or fragment thereof comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO:5.

The examiner does not rely on a reference.

GROUNDS OF REJECTION

Claim 1 stands rejected under 35 U.S.C. § 101 as lacking utility and § 112, first paragraph, for lack of enablement based on the finding of lack of utility. Claim 1 also stands rejected under 35 U.S.C. § 112, first paragraph, as the specification fails to provide an adequate written description of the claimed invention. We affirm the utility and enablement rejections. We reverse the written description rejection.

BACKGROUND

The subject matter of the present appeal is directed to expressed sequence tags (ESTs). See Specification, page 15, lines 9-10. ESTs "are short sequences of randomly selected clones from a cDNA (or complementary DNA) library which are representative of the cDNA inserts of these randomly selected clones." Specification, page 1.

As set forth at page 9, lines 2-4, of appellants' specification "[t]he present invention provides a substantially purified nucleic acid molecule that encodes a maize protein or fragment thereof comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 32236." Of these 32,236 nucleic acid sequences, the originally filed claims where directed to SEQ ID NO: 1 through SEQ ID NO: 4,013. On January 26, 2001 (Paper No. 4), the examiner entered a Restriction requirement into the record, requiring, inter alia, appellants "to elect up to 5 nucleic acid sequences" for consideration on the merits. Paper No. 4, page 3. In response, appellants elected SEQ ID NO:1 through SEQ ID NO:5. The ESTs set forth in SEQ ID NO: 1 through SEQ ID NO:

5 are disclosed to be obtained from cDNA library LIB3115 "generated from maize (RX601, Asgrow Seed Company, Des Moines, Iowa U.S.A.) pooled leaf tissue...." Specification, pages 79-80, Example 1.

The specification sets forth a number of utilities for the nucleic acid molecules of SEQ ID NO: 1 through SEQ ID NO: 5 which are summarized by the examiner (Answer, bridging paragraph, pages 5-6) as follows:

> The specification teaches that the nucleic acids may be used to produce a plant containing reduced levels of a protein (pg. 11), determining an association between a polymorphism and a plant trait (pg. 11), isolating a genetic region or nucleic acid (pg. 11), determining a level or pattern in a plant cell of a protein in a plant (pg. 11), determining a mutation in a plant whose presence is predictive of a mutation affecting a level or pattern of a protein (pg. 13), as molecular tags to isolate genetic regions, isolate genes, map genes, and determine gene function (pg. 14), and identifying tissues (pg. 14)[.] The specification states that the nucleic acid ESTs of the present invention can enable the acquisition of molecular markers, which can be used in breeding schemes, genetic and molecular mapping and cloning of agronomically significant genes (pg. 31).

In the examiner's opinion "[t]hese are non-specific uses that are applicable to nucleic acids in general and not particular or specific to the nucleic acids being claimed." Answer, page 6. For example, the examiner finds (Answer, page 10), "determining whether the claimed nucleic acids have or do not have a polymorphism would require determining whether there was a polymorphism within such a sequence and then determining how to use this information in a patentably meaningful way."[1]

[1] During the Oral Hearing, appellants' representative confirmed that the administrative file contained no evidence that the claimed ESTs where capable of detecting a polymorphism that correlated with any particular trait.

In presenting their case on appeal, appellants focus on use of the claimed nucleic acid molecules to identify the presence or absence of a polymorphism, and their use as probes or as a source for primers. See e.g., Brief, pages 6-12. According to appellants (Brief, page 3), "they have disclosed nucleic acid molecules which, in their current form, provide at least one specific benefit to the public, for example the ability to identify the presence or absence of a polymorphism in a population of maize plants." Furthermore, appellants assert (Brief, page 8), "[t]he specification discloses that the claimed nucleic acid molecules can be used to isolate nucleic acid molecules of other plants and organisms...."

CLAIM CONSTRUCTION

As set forth above, claim 1 on appeal is drawn to a substantially purified nucleic acid molecule that encodes a maize protein or fragment thereof comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO:5. According to appellants' specification (page 15, lines 19-25), the term "substantially purified"

> refers to a molecule separated from substantially all other molecules normally associated with it in its native state. More preferably a substantially purified molecule is the predominant species present in a preparation. A substantially purified molecule may be greater than 60% free, preferably 75% free, more preferably 90% free, and most preferably 95% free from the other molecules (exclusive of solvent) present in the natural mixture. The term "substantially purified" is not intended to encompass molecules present in their native state.

As we understand the claimed invention the use of the transitional term "comprising" does not allow for internal alterations (e.g. insertions or deletions) of

the nucleotide sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 5, but instead only allows for the addition of nucleotides or other molecules at either end of the nucleotide sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 5.[2] In this regard, we recognize, as does the examiner (Answer, page 14), the claim as written encompasses, inter alia, genes, full open reading frames, fusion constructs, and cDNAs.

Accordingly, for the purposes of our review, we interpret the claimed invention as drawn to a nucleic acid molecule, separated from substantially all other molecules normally associated with it in its native state, selected from the group consisting of the nucleic acid molecule defined by the 429 nucleotide sequence set forth in SEQ ID NO: 1, the 413 nucleotide sequence set forth in SEQ ID NO: 2, the 365 nucleotide sequence set forth in SEQ ID NO: 3, the 414 nucleotide sequence set forth in SEQ ID NO: 4, and the 333 nucleotide sequence set forth in SEQ ID NO: 5, with or without any preceding or trailing nucleotides, or other molecules.

DISCUSSION

Utility

The starting point for determining whether a nucleic acid molecule selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 5

[2] This interpretation of the claimed invention was confirmed by appellants' representative during the Oral Hearing.

possesses utility under 35 U.S.C. § 101 is Brenner v. Manson, 383 U.S. 519, 148 USPQ 689 (1966). As set forth in Brenner, at 534-35, 148 USPQ at 695[3],

> the basic quid pro quo contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with substantial utility. Unless and until [an invention] is refined and developed to this point--where specific benefit exists in currently available form--there is insufficient justification for permitting an applicant to engross what may prove to be a broad field.

In considering the issues presented in this appeal, special attention must be paid to the Brenner court's statement that a patent should issue only when an invention possesses "substantial utility," i.e., "where a specific benefit exists in currently available form." Whether a claimed invention is useful under 35 U.S.C. § 101 is a question of fact. Cross v. Iizuka, 753 F.2d 1040, 1044 n.7, 224 USPQ 739, 742 n.7 (Fed. Cir. 1985).

At issue in Brenner was a claim to "a chemical process which yields an already known product whose utility—other than as a possible object of scientific inquiry—ha[d] not yet been evidenced." Id. at 529, 148 USPQ at 693. The Patent Office had rejected the claimed process for lack of utility, on the basis that the product produced by the claimed process had not been shown to be useful. See id. at 521-22, 148 USPQ at 690. On appeal, the Court of Customs and Patent Appeals reversed, on the basis that "where a claimed process produces a

---

[3] In discussing the issue of utility under 35 U.S.C. § 101, the Federal Circuit and the Court of Customs and Patent Appeals since Brenner, have used the phrases "substantial utility" and "practical utility" interchangeably. See e.g., Fujikawa v. Wattanasin, 93 F.3d 1559, 1963-1964, 39 USPQ2d 1895, 1898-1899 (Fed. Cir. 1996) ("It is well established that a patent may not be granted to an invention unless substantial or practical utility for the invention has been discovered and disclosed.").

known product it is not necessary to show utility for the product." Id. at 522, 148 USPQ at 691.

The Brenner Court noted that although § 101 requires that an invention be "useful," that "simple, everyday word can be pregnant with ambiguity when applied to the facts of life." Id. at 529, 148 USPQ at 693. Thus,

> [i]t is not remarkable that differences arise as to how the test of usefulness is to be applied to chemical processes. Even if we knew precisely what Congress meant in 1790 when it devised the "new and useful" phraseology and in subsequent re-enactments of the test, we should have difficulty in applying it in the context of contemporary chemistry, where research is as comprehensive as man's grasp and where little or nothing is wholly beyond the pale of "utility"—if that word is given its broadest reach.

Id. at 530, 148 USPQ at 694.[4]

The Court, finding "no specific assistance in the legislative materials underlying § 101," based its analysis on "the general intent of Congress, the purposes of the patent system, and the implications of a decision one way or the other." Id. at 532, 148 USPQ at 695. The Court concluded that "[t]he basic quid pro quo contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with substantial utility. Unless and until a process is refined and developed to this point—where specific benefit exists in currently available form—there is insufficient justification for permitting an applicant to engross what may prove to be a broad field." Id. at 534-35, 148 USPQ at 695.

---

[4] The invention at issue in Brenner was a process, but the Court expressly noted that its holding "would apply equally to the patenting of the product produced by the process." Id. at 535, 148 USPQ at 695-96.

The Court considered and rejected the applicant's argument that attenuating the requirement of utility "would encourage inventors of new processes to publicize the event for the benefit of the entire scientific community, thus widening the search for uses and increasing the fund of scientific knowledge." The Court noted that, while there is value to encouraging disclosure, "a more compelling consideration is that a process patent in the chemical field, which has not been developed and pointed to the degree of specific utility, creates a monopoly of knowledge which should be granted only if clearly commanded by the statute. Until the process claim has been reduced to production of a product shown to be useful, the metes and bounds of that monopoly are not capable of precise delineation. It may engross a vast, unknown, and perhaps unknowable area. Such a patent may confer power to block off whole areas of scientific development." Id. at 534, 148 USPQ at 695.

The Court took pains to note that it did not "mean to disparage the importance of contributions to the fund of scientific information short of the invention of something 'useful,'" and that it was not "blind to the prospect that what now seems without 'use' may tomorrow command the grateful attention of the public." Id. at 535-36, 148 USPQ at 696. Those considerations did not sway the Court, however, because "a patent is not a hunting license. It is not a reward for the search, but compensation for its successful conclusion." Id.

Subsequent decisions of the CCPA and the Court of Appeals for the Federal Circuit have added further layers of judicial gloss to the meaning of

§ 101's utility requirement. The first opinion of the CCPA applying Brenner was In re Kirk, 376 F.2d 936, 153 USPQ 48 (CCPA 1967). The invention claimed in Kirk was a set of steroid derivatives said to have valuable biological properties and to be of value "in the furtherance of steroidal research and in the application of steroidal materials to veterinary or medical practice." Id. at 938, 153 USPQ at 50. The claims had been rejected for lack of utility. In response, the applicants submitted an affidavit which purportedly "show[ed] that one skilled in the art would be able to determine the biological uses of the claimed compounds by routine tests." Id. at 939, 153 USPQ at 51.

The court held that "nebulous expressions [like] 'biological activity' or 'biological properties'" did not adequately convey how to use the claimed compounds. Id. at 941, 153 USPQ at 52. Nor did the applicants' affidavit help their case: "the sum and substance of the affidavit appear[ed] to be that one of ordinary skill in the art would know 'how to use' the compounds to find out in the first instance whether the compounds are—or are not—in fact useful or possess useful properties, and to ascertain what those properties are." Id. at 942, 153 USPQ at 53.

The Kirk court held that an earlier CCPA decision, holding that a chemical compound meets the requirements of § 101 if it is useful to chemists doing research on steroids, had effectively been overruled by Brenner. "There can be no doubt that the insubstantial, superficial nature of vague, general disclosures or arguments of 'useful in research' or 'useful as building blocks of value to the

researcher' was recognized, and clearly rejected, by the Supreme Court" in Brenner. See Kirk, 376 F.2d at 945, 153 USPQ at 55.

More recently, in In re Ziegler, 992 F.2d 1197, 26 USPQ2d 1600 (Fed. Cir. 1993), the Federal Circuit considered the degree of specificity required to show utility for a claim to polypropylene. The U.S. application on appeal in Ziegler claimed priority to a German application filed in 1954. "In the German application, Ziegler disclosed only that solid granules of polypropylene could be pressed into a flexible film with a characteristic infrared spectrum and that the polypropylene was 'plastic-like.'" Id. at 1203, 26 USPQ2d at 1605. "Ziegler did not assert any practical use for the polypropylene or its film, and Ziegler did not disclose any characteristics of the polypropylene or its film that demonstrated its utility." Id. The court held that the German application did not satisfy the requirements of § 101 and therefore could not be relied on to overcome a rejection based on an intervening reference. Id. "[At] best, Ziegler was on the way to discovering a practical utility for polypropylene at the time of the filing of the German application; but in that application Ziegler had not yet gotten there." Id.

On the other hand, the CCPA reversed a rejection for lack of utility in In re Jolles, 628 F.2d 1322, 206 USPQ 885 (CCPA 1980). The applicant in Jolles claimed pharmaceutical compositions that were disclosed to be useful in treating acute myeloblastic leukemia. See id. at 1323, 206 USPQ at 886. The active ingredients in the compositions were closely related to daunorubicin and doxorubicin, both of which were "well recognized in the art as valuable for use in

cancer chemotherapy." Id., 206 USPQ at 887. The applicant also submitted declaratory evidence showing that eight of the claimed compositions were effective in treating tumors in a mouse model, and one was effective in treating humans. See id. at 1323-24, 206 USPQ at 887-88. The court noted that the data derived from the mouse model were "relevant to the treatment of humans and [were] not to be disregarded," id. at 1327, 206 USPQ at 890, and held that the evidence was sufficient to support the asserted therapeutic utility. See id. at 1327-28, 206 USPQ at 891.

The Federal Circuit held in Cross v. Iizuka, 753 F.2d 1040, 224 USPQ 739 (Fed. Cir. 1985), that in vivo testing (as in Jolles) was not necessarily required to show utility in the pharmaceutical context. The Cross court stated that "[it] is axiomatic that an invention cannot be considered 'useful,' in the sense that a patent can be granted on it, unless substantial or practical utility for the invention has been discovered and disclosed where such utility would not be obvious." Id. at 1044, 224 USPQ at 742 (citing Brenner v. Manson). The court "perceive[d] no insurmountable difficulty, under appropriate circumstances, in finding that the first link in the screening chain, in vitro testing, may establish a practical utility for the compound in question." Id. at 1051, 224 USPQ at 748. Successful in vitro testing could provide an immediate benefit to the public, by "marshal[ling] resources and direct[ing] the expenditure of effort to further in vivo testing of the most potent compounds ..., analogous to the benefit provided by the showing of an in vivo utility." Id. On the facts of that case – successful in vitro testing supplemented by similar in vitro and in vivo activities of structurally similar

compounds – the court held that in vitro activity was sufficient to meet the requirements of § 101. See id.

The Federal Circuit confirmed in In re Brana, 51 F.3d 1560, 34 USPQ2d 1436 (Fed. Cir. 1995), that human testing is not necessary to establish utility for a method of treatment. The invention claimed in Brana was a group of compounds disclosed to have antitumor activity. See id. at 1562, 34 USPQ2d at 1437-38. The specification disclosed that the claimed compounds had higher antitumor activity than related compounds known to have antitumor activity, and the applicants provided declaratory evidence of in vivo activity against tumors in a mouse model. See id., 34 USPQ2d at 1438. The court held that these data were sufficient to satisfy § 101; usefulness in patent law does not require that the invention be ready to be administered to humans. See id. at 1567, 34 USPQ2d at 1442.

Several lessons can be drawn from Brenner and its progeny. First, § 101's requirement that an invention be "useful" is not to be given its broadest reach, such that little or nothing of a chemical nature would be found to lack utility. See Brenner, 383 U.S. at 530, 148 USPQ at 694. Thus, not every "use" that can be asserted will be sufficient to satisfy § 101. For example, the steroid compound at issue in Brenner was useful as a possible object of scientific inquiry, and the polypropylene claimed in Ziegler was useful for pressing into a flexible film, yet both lacked sufficient utility to satisfy § 101. See Brenner, 383 U.S. at 529, 148 USPQ at 696; Ziegler, 992 F.2d at 1203, 26 USPQ2d at 1605.

Rather than setting a de minimis standard, § 101 requires a utility that is "substantial", i.e., one that provides a specific benefit in currently available form. Brenner, 383 U.S. at 534-35, 148 USPQ at 695. This standard has been found to be met by pharmaceutical compositions shown to be useful in mouse models and in humans for treating acute myeloblastic leukemia (Jolles, 628 F.2d at 1327-28, 206 USPQ at 891); by evidence showing successful in vitro testing supplemented by similar in vitro and in vivo activities of structurally similar compounds (Cross, 753 F.2d at 1051, 224 USPQ at 748); and by evidence showing in vivo antitumor activity in mice, combined with a disclosure that the claimed compounds had higher antitumor activity than a related compound known to have antitumor activity (Brana, 51 F.3d at 1567, 34 USPQ2d at 1442).

By contrast, Brenner's standard has been interpreted to mean that "vague, general disclosures or arguments of 'useful in research' or 'useful as building blocks of value to the researcher'" would not satisfy § 101. See Kirk, 376 F.2d at 945, 153 USPQ at 55 (interpreting Brenner). Likewise, a disclosure of a "plastic-like" polypropylene capable of being pressed into a flexible film was held to show that the applicant was "at best ... on the way to discovering a practical utility for polypropylene at the time of the filing," but not yet there. Ziegler, at 1203, 26 USPQ2d at 1605.

With these principles in mind we turn to the issues at hand. Of the many utilities asserted in the specification, two have received the most attention in the briefing in this appeal, i.e., identification and detection of polymorphisms and use

as probes or as a source for primers. We shall focus on these asserted utilities first and then address the other arguments set forth in the briefing.

a. Polymorphisms

This utility is discussed at pages 35-42 of the specification in terms of what polymorphisms are and how one would go about determining the existence of a polymorphism. The discussion in this portion of the specification, however, is not specific to the nucleotide molecules depicted in SEQ ID NO: 1 through SEQ ID NO: 5. To the contrary, according to appellants' specification (page 35, lines 25-26), "one or more of the [32,236] EST nucleic acid molecules (or a sub-fragment thereof) may be employed as a marker nucleic acid molecule to identify ... polymorphism(s)." The specification does not explain why any of the 32,236 nucleotide molecules disclosed in the specification, or more specifically the five nucleotide molecules depicted in SEQ ID NO: 1 through SEQ ID NO: 5, would in fact be useful in detecting polymorphisms.

Rather, appellants argue (Brief, page 7), "the claimed nucleic acid molecules have utility even if the absence of a particular polymorphism is detected. Indeed, the absence of a polymorphism usually demonstrates that the two (or more) populations being compared share a common genetic heritage." In other words, appellants' position is that an EST by definition possesses patentable utility because it can be used by itself in determining whether populations share a common genetic heritage. While that may be a "utility," we do not find that it is a substantial utility.

Without knowing any further information in regard to the gene represented by an EST, as here, detection of the presence or absence of a polymorphism provides the barest information in regard to genetic heritage. As the examiner explains (Answer, bridging paragraph, pages 10-11):

> Polymorphisms are natural variations within sequences which themselves may not have any meaningful use. Therefore, determining whether the claimed nucleic acids [(or nucleic acids detected by the claimed nucleic acids)] have or do not have a polymorphism would require determining whether there was a polymorphism within such a sequence and then determining how to use this information in a patentably meaningful way. The [a]ppellant also argues, "many of these uses are directly analogous to a microscope". This argument has been reviewed but is not convincing because the microscope provides information to the scientist which is automatically useful. For example, the microscope may be used for identification and differentiation between gram-positive and gram-negative bacteria. The differentiation of bacteria facilitates in the administration of proper antibiotics. For example, if the microscope is used to determine whether Staph is present or whether Strep is present provides valuable information to the scientist and/or doctor for treating patients. The instant invention, however, provides no information to this extent. If the scientist determines that SEQ ID NO: 1 is present, the scientist does not know how to use this information. Thus, the identification of SEQ ID NO: 1 is not a substantial utility.

In contrast, at the other end of the "utility spectrum" would be information gleaned from detecting the presence or absence of a polymorphism when it is known what effect the gene from which the EST is derived has in the development and/or phenotype of the plant. Somewhere between having no knowledge (the present circumstances) and having complete knowledge of the gene and its role in the plant's development and/or phenotype lies the line between "utility" and "substantial utility." We need not draw the line or further

define it in this case because the facts in this case represent the lowest end of the spectrum, i.e., an insubstantial use.

b. Probes or source of primers

Appellants argue that the "specification discloses that the claimed nucleic acid molecules can be used to isolate nucleic acid molecules of other plants and organisms...." Appeal Brief, page 8. While that may be true, it begs the question of what substantial use such nucleic acid molecules would have? Again, the present specification does not attribute any property in terms of plant trait, or phenotype to any of the nucleotide molecules set forth in SEQ ID NO: 1 through SEQ ID NO: 5. In the absence of such information, using the claimed molecules to isolate other molecules, which themselves lack substantial utility, does not represent a substantial utility.

Appellants also assert that the claimed nucleic acid molecules may be used in a "chromosome walk." Brief, pages 8-9. According to appellants (Brief, page 9),

> The claimed nucleic acid molecules provide a particularly appropriate and demonstrably useful starting point for a walk to isolate a promoter that is active in leaves at the time of anthesis. Isolation of such a promoter would be desirable and particularly useful because it allows expression of proteins at that important developmental state, including proteins that provide disease resistance. Because the claimed nucleic acid molecules were isolated from leaves, they provide an appropriate starting point for isolating a promoter active in leaves. A random nucleic acid molecule does not provide an equally good starting point to isolate such a promoter....

As we understand this argument, the claimed ESTs may be useful in searching for promoters that are only active in leaves at the time of anthesis. The

specification, however, fails to demonstrate that any of the nucleic acid molecules set forth in SEQ ID NO: 1 through SEQ ID NO: 5 would be useful in obtaining a successful result from such a search. As set forth at page 34, lines 14-19 of appellants' specification,

> The [32,236] nucleic acid molecules of the present invention may be used to isolate promoters of tissue enhanced[,] tissue specific, cell-specific, cell –type, developmentally or environmentally regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences of these genes from genomic libraries, for example, using genomic screening methods and PCR techniques would result in the isolation of useful promoters and transcriptional regulatory elements.

The specification does not provide any expectation of successfully using any of the 32,236 nucleic acid molecules disclosed in the specification, or more specifically the five nucleic acid molecules depicted in SEQ ID NO: 1 through SEQ ID NO: 5, to isolate promoters of tissue enhanced, tissue specific, cell-specific, cell–type, developmentally or environmentally regulated expression profiles.

Furthermore, notwithstanding appellants' assertion (Brief, page 9), there is no evidence on this record that any of the nucleic acid molecules depicted in SEQ ID NO: 1 through SEQ ID NO: 5 are tissue or cell-type specific, or developmentally or environmentally regulated. In this regard, we note that the claimed nucleic acid molecules were isolated from the cDNA library LIB3115. Specification, page 80, lines 5-6. There is no evidence on this record that LIB3115 is a subtractive cDNA library, wherein nucleic acid molecules from other maize tissue, or from other developmental stages, was subtracted (removed)

from the library. Compare, for example, the subtractive cDNA library LIB3153 which is disclosed (specification, page 83, lines 17-19) to be "generated by subtracting driver cDNA, which is prepared from kernels harvested from 15 DAP [days after pollination] maize plants, from target cDNA, which is prepared from endosperms harvested from 5-8 day[s] after pollination (DAP) maize plants." In contrast to the claimed nucleic acid molecules, nucleic acid molecules SEQ ID NO: 24,931 through SEQ ID NO: 25,680 are from the subtractive cDNA library LIB3153.

In our opinion, the claimed nucleic acid molecules having the sequences identified as SEQ ID NO: 1 through SEQ ID NO: 5, represent five randomly selected nucleic acid molecules isolated from pooled leaf tissue at the time of anthesis. Notwithstanding appellants' emphasis on "anthesis," for the foregoing reasons, we find no evidence on this record that any of appellants' five randomly selected nucleic acid molecules are expressed only at the time of "anthesis." Accordingly, despite appellants' assertion to the contrary, there is no reasonable expectation that any of the claimed nucleic acid molecules would be capable of isolating a promoter that was only active in leaves at the time of anthesis. As appellants recognize (Brief, page 9), "[a] random nucleic acid molecule does not provide an equally good starting point to isolate such a promoter" compared to a nucleic acid molecule that is known to be specifically associated with this stage of plant development.

We recognize appellants' argument (Brief, bridging sentence, pages 9-10), "[a]n invention may be 'less effective than existing devices but nevertheless

meet the statutory criteria for patentability.' Custom Accessories, Inc. v. Jeffrey-Allan Indus., 807 F.2d 955, 960 n.12, 1 U.S.P.Q.2d 1196, 1199 n.12 (Fed. Cir. 1986)." While we agree with appellants' statement, we fail to see how it applies to appellants' claimed invention, wherein there is no evidence or expectation that the claimed nucleic acid molecules would be "effective" at all. In this regard, we remind appellants that an invention does not have utility sufficient to satisfy § 101 until it is "refined and developed" to the point of providing a specific benefit in currently available form. See, e.g., Brenner, 383 U.S. at 534, 148 USPQ at 695.

An invention certainly can have a utility that is shared by other compounds or compositions. Take, for example, an application that claims ibuprofen and discloses that it is useful as an analgesic. No one would argue that a claim to ibuprofen lacks utility simply because aspirin and acetaminophen are also useful as analgesics. On the other hand, not every utility will satisfy § 101, even if the utility is shared by a class of inventions. Assume that the above-described application did not disclose that ibuprofen was an analgesic but only disclosed that it is useful because it can be used to fill a jar, which would then be useful as a paperweight. There would be little doubt that this disclosed utility would not satisfy § 101, even though the utility is shared by a large class of inventions, viz., those whose physical embodiments have mass. So while a utility need not be unique to a claimed invention, it must nonetheless be specific, and in currently available form, in order to satisfy § 101.

c. Other Arguments

Appellants argue that the specification "discloses additional utilities for the claimed nucleic acid molecules including introduction of the claimed nucleic acid molecules into a plant or plant cell (either as sense or antisense inhibitors), which can then be used to screen for compounds such as a herbicide." Brief, page 6. Specifically, appellants argue (id.) that "a compound can be provided to both an antisense plant and a control plant (no antisense) and the effect of the compound on the plant can be monitored." Appellants analogize this proposed procedure to a "cell-based assay" which appellants assert to have a "legally sufficient utility." Id.

Suffice it to say that an otherwise uncharacterized nucleic acid molecule is being claimed in this application, not an assay. The portion of the specification cited in support of this argument (page 73, line 17 through page 74, line 17) indicates that the nucleic acid molecule must be introduced into a plant cell and transcribed using an appropriate promoter to result in the suppression of an endogenous protein. The specification does not indicate that such a method is feasible when the nucleic acid to be used is uncharacterized[5] as here. Such a use does not provide a specific or substantial benefit in currently available form.

Appellants also argue that the claimed nucleic acids are useful to measure the level of mRNA in a sample through use of microarray technology

[5] To emphasize the uncharacterized nature of appellants' invention we note the examiner's finding (Answer, page 17) that translating SEQ ID NO: 5 in all 6 possible reading frames reveals that the sequence contains numerous stop codons which would terminate the translation of a protein, or protein fragment, encoded thereby.

and use as molecular markers. Brief, page 6. In regard to microarrays, appellants argue (id. fn. 3) that it is "standard practice" to screen populations of nucleic acids with EST sequences without characterizing each and every target mRNA. We find that the asserted utility of the claimed nucleic acid—as one component of an assay for monitoring gene expression—does not satisfy the utility requirement of § 101. Such a use does not provide a specific benefit in currently available form. We accept, for argument's sake, that a person skilled in the art could use the claimed nucleic acid, in combination with other nucleic acids, to monitor changes in expression of the gene that encompasses the nucleic acid depicted in e.g., SEQ ID NO: 1. However, the specification provides no guidance that would allow a skilled artisan to use data relating to expression of such a gene in any practical way. The specification simply provides no guidance regarding what the SEQ ID NO: 1-specific information derived from a gene expression experiment would mean. As the examiner points out (Answer, page 9), "the instant claimed nucleic acids appear to require further experimentation on the material itself to determine the function and properties of the claimed nucleic acids."

To highlight the examiner's assertion, suppose, for example, that a researcher found that SEQ ID NO: 1 expression was increased when a cell was treated with a particular agent. The specification provides no basis on which a skilled worker would be able to determine whether that result is meaningful. Maybe the meaning in a change in SEQ ID NO: 1 expression would depend on other factors, but again the specification provides no hint as to what other factors

might be important. Would it depend on what agent is used, what cell type is used, the behavior of other genes (if so, which genes and what behavior is significant), the degree of increase? The specification simply provides no guidance as to how to interpret the results that might be seen using SEQ ID NO: 1 in a gene expression assay.

In effect, appellants' position is that the claimed nucleic acids are useful because those of skill in the art could experiment with them and figure out for themselves what any observed experimental results might mean. We do not agree that such a disclosure provides a "specific benefit in currently available form." Rather, the present case seems analogous to Brenner. In Brenner, the applicant claimed a method of making a compound but disclosed no utility for the compound. 383 U.S. at 529, 148 USPQ at 693. The Court held that a process lacks utility if it produces a product that lacks utility. Id. at 534, 148 USPQ at 695. Here, appellants claim a product asserted to be useful in a method of generating gene-expression data, but the specification does not disclose how to interpret those data. Just as the process claimed in Brenner lacked utility because the specification did not disclose how to use the end-product, the products claimed here lack utility, because even if used in gene expression assays, the specification does not disclose how to use SEQ ID NO: 1-specific gene expression data.

Assuming arguendo, that a generic gene expression assay—one based on monitoring expression of thousands of uncharacterized nucleic acids would provide a useful tool for, e.g., drug discovery, it does not follow that each one of

the nucleic acids represented in the assay individually has patentable utility. Although each nucleic acid in the assay contributes to the data generated by the assay overall, the contribution of a single nucleic acid—its data point—is only a tiny contribution to the overall picture. The Brenner Court held that § 101 sets more than a de minimis standard for utility. Therefore, the patentable utility of a gene expression assay, for example, does not necessarily mean that each tiny component of the assay also has patentable utility. A patentable utility divided by a thousand does not necessarily equal a thousand patentable utilities. Each claimed invention must be shown to meet § 101's utility requirement in order to be patentable; it must provide a specific benefit in currently available form. Providing a single data point among thousands or millions, even if the thousands or millions of data points collectively are useful, does not meet this standard. The Supreme Court noted that the patent system contemplates a basic quid pro quo: in exchange for the legal right to exclude others from his invention for a period of time, an inventor discloses his invention to the public. See Brenner, 383 U.S. at 534, 148 USPQ at 695. The Brenner Court held that the grant of patent rights to an applicant is justified only by disclosure of an invention with substantial utility – a specific benefit in currently available form. Until the invention has been refined and developed to this point, the Court held, the applicant has not met his side of the bargain, and has not provided a disclosure sufficient to justify a grant of the right to exclude others. See id.

We reach the same conclusion in regard to appellants' assertion that the nucleic acid molecules depicted in SEQ ID NO: 1 through SEQ ID NO: 5 are

useful as a molecular marker or probe. It is not seen that the one data point which may be provided by using the uncharacterized nucleic acid molecule of SEQ ID NO: 1 as a molecular marker or probe represents a substantial use.

Appellants argue that ESTs have real world value as seen from the "growth of a multi-million dollar industry in the United States premised on the usefulness of ESTs." Brief, page 11. Since appellants fail to provide any suggestion on which use of ESTs this industry is premised on, we can only assume that appellants are referring to the potential usefulness of EST databases, clone sets or microarrays. Suffice it to say, the claims on appeal are not directed to EST databases, clone sets and/or microarrays. Again, it is not seen that the one data point which may be provided by using the uncharacterized nucleic acid molecules of SEQ ID NO: 1 through SEQ ID NO: 5 in such devices represents a substantial use.

For the foregoing reasons we affirm the rejection of claim 1 under 35 U.S.C. § 101.

Enablement

According to the examiner (Answer, page 13, emphasis removed), "since the claimed invention is not supported by either a specific, substantial asserted utility or a well established utility for the reasons set forth [in support of the rejection under 35 U.S.C. § 101] one skilled in the art clearly would not know how to use the claimed invention." This rejection is simply a corollary of the finding of lack of utility. Appellants assert (Brief, page 12), this rejection should be reversed for the same reasons set forth in their arguments regarding the

rejection under 35 U.S.C. § 101. Thus, our conclusion with respect to the § 101 issue will also apply to this aspect of the § 112 (enablement) issue. On this basis we affirm the rejection of claim 1 under the enablement provision of 35 U.S.C. § 112, first paragraph.

Written description

This rejection stands on a different footing. As we understand the examiner's argument the use of the transitional phrase "comprising" in appellants' claimed invention results in appellants claiming a large genus of nucleic acid molecules which are not adequately described by SEQ ID NO: 1 through SEQ ID NO: 5. Answer, pages 13-16. Apparently the examiner is of the opinion that the claimed invention should be limited to nucleic acid molecules as set forth in SEQ ID NO: 1 through SEQ ID NO: 5. In response appellants argue (Brief, page 14, original footnote omitted),

> Applicants have provided the nucleotide sequences required by the claims, i.e., SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, and SEQ ID NO: 5, and have thus established possession of the claimed invention. The fact that the claims at issue are intended to cover molecules that include the recited sequences joined with additional sequences[6] does not mean that [a]pplicants were any less in possession of the claimed nucleic acid molecules.

As discussed supra, as we understand the claimed invention, the use of the transitional term "comprising" does not allow for internal alterations (e.g. insertions or deletions) of the nucleotide sequences set forth in SEQ ID NO: 1

[6] By way of examples appellants explain (Brief, bridging paragraph, pages 14-15) that the specification discloses, inter alia, the claimed nucleic acid molecules joined together with vectors, and other nucleic acids (e.g. fusion nucleic acid molecules) and detectable labels.

through SEQ ID NO: 5, but instead only allows for the addition of nucleotides or other molecules at either end of the nucleotide sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 5. We agree with appellants that they have provided an adequate written description of nucleic acid molecules with the sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 5. That the claimed nucleic acid molecules may have other molecules attached to either, or both of their 5' or 3' ends does not diminish appellants' adequate written description of the nucleic acids molecules with the sequences set forth in SEQ ID NO: 1 through SEQ ID NO: 5 as claimed.

Accordingly, we reverse the rejection of claim 1 under the written description provision of 35 U.S.C. § 112, first paragraph.

No time period for taking any subsequent action in connection with this appeal may be extended under 37 CFR § 1.136(a).

AFFIRMED

William F. Smith
Administrative Patent Judge

Donald E. Adams
Administrative Patent Judge

Eric Grimes
Administrative Patent Judge

) BOARD OF PATENT
) APPEALS AND
) INTERFERENCES

Lawrence M Lavin Jr Esq
Monsanto Company
Patent Department E2NA
800 N Lindbergh Boulevard
St. Louis MO 63167

NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH PLANTS

Field of the Invention

The present invention is in the field of plant biochemistry. More specifically the invention relates to nucleic acid molecules that encode proteins and fragments of proteins produced in plant cells, in particular, maize plants. The invention also relates to proteins and fragments of proteins so encoded and antibodies capable of binding the proteins. The invention also relates to methods of using the nucleic acid molecules, proteins and fragments of proteins.

Background of the Invention

I. EXPRESSED SEQUENCE TAG NUCLEIC ACID MOLECULES

Expressed sequence tags, or ESTs, are short sequences of randomly selected clones from a cDNA (or complementary DNA) library which are representative of the cDNA inserts of these randomly selected clones. McCombie, *et al., Nature Genetics, 1:*124-130 (1992); Kurata *et al., Nature Genetics, 8:* 365-372 (1994); Okubo, *et al., Nature Genetics, 2:* 173-179 (1992), all of which references are incorporated herein in their entirety.

Using conventional methodologies, cDNA libraries can be constructed from the mRNA (messenger RNA) of a given tissue or organism using poly dT primers and reverse transcriptase (Efstratiadis, *et al., Cell 7*:279-288 (1976), the entirety of which is herein incorporated by reference; Higuchi, *et al., Proc. Natl. Acad. Sci. (U.S.A.) 73*:3146-3150 (1976), the entirety of which is herein incorporated by reference; Maniatis, *et al., Cell 8*:163 (1976) the entirety of which is herein incorporated by reference; Land, *et al., Nucleic Acids Res. 9*:2251-2266 (1981), the entirety of which is herein incorporated by reference; Okayama, *et al., Mol. Cell. Biol. 2*:161-170 (1982), the entirety of which is herein incorporated by reference; Gubler, *et al., Gene 25*:263 (1983), the entirety of which is herein incorporated by reference).

Several methods may be employed to obtain full-length cDNA constructs. For example, terminal transferase can be used to add homopolymeric tails of dC residues to the free 3'

hydroxyl groups (Land, *et al.*, *Nucleic Acids Res. 9*:2251-2266 (1981), the entirety of which is herein incorporated by reference). This tail can then be hybridized by a poly dG oligo which can act as a primer for the synthesis of full length second strand cDNA. Okayama and Berg, report a method for obtaining full length cDNA constructs. This method has been simplified by using synthetic primer-adapters that have both homopolymeric tails for priming the synthesis of the first and second strands and restriction sites for cloning into plasmids (Coleclough, *et al.*, *Gene 34*:305-314 (1985), the entirety of which is herein incorporated by reference) and bacteriophage vectors (Krawinkel, *et al.*, *Nucleic Acids Res. 14*:1913 (1986), the entirety of which is herein incorporated by reference; and Han, *et al.*, *Nucleic Acids Res. 15*:6304 (1987), the entirety of which is herein incorporated by reference).

These strategies have been coupled with additional strategies for isolating rare mRNA populations. For example, a typical mammalian cell contains between 10,000 and 30,000 different mRNA sequences. Davidson, *Gene Activity in Early Development*, 2nd ed., Academic Press, New York (1976). The number of clones required to achieve a given probability that a low-abundance mRNA will be present in a cDNA library is $N = (\ln(1-P))/(\ln(1-1/n))$ where N is the number of clones required, P is the probability desired, and 1/n is the fractional proportion of the total mRNA that is represented by a single rare mRNA. (Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press (1989), the entirety of which is herein incorporated by reference.).

A method to enrich preparations of mRNA for sequences of interest is to fractionate by size. One such method is to fractionate by electrophoresis through an agarose gel (Pennica, *et al.*, *Nature 301*:214-221 (1983), the entirety of which is herein incorporated by reference). Another such method employs sucrose gradient centrifugation in the presence of an agent, such as methylmercuric hydroxide, that denatures secondary structure in RNA (Schweinfest, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 79*:4997-5000 (1982), the entirety of which is herein incorporated by reference).

A frequently adopted method is to construct equalized or normalized cDNA libraries (Ko, *Nucleic Acids Res. 18*:5705-5711 (1990), the entirety of which is herein incorporated by reference; Patanjali, S. R. *et al., Proc. Natl. Acad. Sci. (U.S.A.) 88*:1943-1947 (1991), the entirety of which is herein incorporated by reference). Typically, the cDNA population is normalized by subtractive hybridization. Schmid, *et al., J. Neurochem. 48*:307-312 (1987) the entirety of which is herein incorporated by reference; Fargnoli *et al., Anal. Biochem. 187*:364-373 (1990) the entirety of which is herein incorporated by reference; Travis, *et al., Proc. Natl. Acad. Sci (U.S.A.) 85*:1696-1700 (1988) the entirety of which is herein incorporated by reference; Kato, *Eur. J. Neurosci. 2*:704 (1990); and Schweinfest, *et al., Genet. Anal. Tech. Appl. 7*:64 (1990), the entirety of which is herein incorporated by reference). Subtraction represents another method for reducing the population of certain sequences in the cDNA library. Swaroop, *et al., Nucleic Acids Res. 19*:1954 (1991), the entirety of which is herein incorporated by reference).

ESTs can be sequenced by a number of methods. Two basic methods may be used for DNA sequencing, the chain termination method of Sanger *et al., Proc. Natl. Acad. Sci. (U.S.A.) 74:* 5463-5467 (1977), the entirety of which is herein incorporated by reference and the chemical degradation method of Maxam and Gilbert, *Proc. Nat. Acad. Sci. (U.S.A.) 74:* 560-564 (1977), the entirety of which is herein incorporated by reference. Automation and advances in technology such as the replacement of radioisotopes with fluorescence-based sequencing have reduced the effort required to sequence DNA (Craxton, *Methods, 2:* 20-26 (1991), the entirety of which is herein incorporated by reference; Ju *et al., Proc. Natl. Acad. Sci. (U.S.A.) 92:* 4347-4351 (1995), the entirety of which is herein incorporated by reference; Tabor and Richardson, *Proc. Natl. Acad. Sci. (U.S.A.) 92:* 6339-6343 (1995), the entirety of which is herein incorporated by reference). Automated sequencers are available from, for example, Pharmacia Biotech, Inc., Piscataway, New Jersey (Pharmacia ALF), LI-COR, Inc., Lincoln, Nebraska (LI-COR 4,000) and Millipore, Bedford, Massachusetts (Millipore BaseStation).

In addition, advances in capillary gel electrophoresis have also reduced the effort required to sequence DNA and such advances provide a rapid high resolution approach for sequencing

DNA samples (Swerdlow and Gesteland, *Nucleic Acids Res. 18*:1415-1419 (1990); Smith, *Nature 349*:812-813 (1991); Luckey *et al.*, *Methods Enzymol. 218*:154-172 (1993); Lu *et al.*, *J. Chromatog. A. 680*:497-501 (1994); Carson *et al.*, *Anal. Chem. 65*:3219-3226 (1993); Huang *et al.*, *Anal. Chem. 64*:2149-2154 (1992); Kheterpal *et al.*, *Electrophoresis 17*:1852-1859 (1996); Quesada and Zhang, *Electrophoresis 17*:1841-1851 (1996); Baba, *Yakugaku Zasshi 117*:265-281 (1997), all of which are herein incorporated by reference in their entirety).

ESTs longer than 150 bases have been found to be useful for similarity searches and mapping. (Adams, *et al.*, *Science 252*:1651-1656 (1991), herein incorporated by reference.) EST sequences normally range from 150-450 bases. This is the length of sequence information that is routinely and reliably generated using single run sequence data. Typically, only single run sequence data is obtained from the cDNA library, Adams, *et al.*, *Science 252*:1651-1656 (1991). Automated single run sequencing typically results in an approximately 2-3% error or base ambiguity rate. (Boguski, *et al.*, *Nature Genetics, 4*:332-333 (1993), the entirety of which is herein incorporated by reference).

EST databases have been constructed or partially constructed from, for example, *C. elegans* (McCombrie, *et al.*, *Nature Genetics 1*:124-131 (1992), human liver cell line HepG2 (Okubo, *et al.*, *Nature Genetics 2*:173-179 (1992)), human brain RNA (Adams, *et al.*, *Science 252*:1651-1656 (1991); Adams, *et al.*, *Nature 355*:632-635 (1992)), *Arabidopsis*, (Newman, *et al.*, *Plant Physiol.* 106:1241-1255 (1994)); and rice (Kurata, *et al.*, *Nature Genetics 8*:365-372 (1994).

II. SEQUENCE COMPARISONS

A characteristic feature of a protein or DNA sequence is that it can be compared with other known protein or DNA sequences. Sequence comparisons can be undertaken by determining the similarity of the test or query sequence with sequences in publicly available or propriety databases ("similarity analysis") or by searching for certain motifs ("intrinsic sequence analysis")(*e.g. cis* elements)(Coulson, *Trends in Biotechnology, 12:* 76-80 (1994), the entirety of

which is herein incorporated by reference; Birren, *et al., Genome Analysis, 1*: 543-559 (1997), the entirety of which is herein incorporated by reference).

Similarity analysis includes database search and alignment. Examples of public databases include the DNA Database of Japan (DDBJ)(http://www.ddbj.nig.ac.jp/); Genebank (http://www.ncbi.nlm.nih.gov/web/Genbank/Index.htlm); and the European Molecular Biology Laboratory Nucleic Acid Sequence Database (EMBL) (http://www.ebi.ac.uk/ebi_docs/embl_db.html). A number of different search algorithms have been developed, one example of which are the suite of programs referred to as BLAST programs. There are five implementations of BLAST, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology, 12:* 76-80 (1994); Birren, *et al., Genome Analysis, 1:* 543-559 (1997)).

BLASTN takes a nucleotide sequence (the query sequence) and its reverse complement and searches them against a nucleotide sequence database. BLASTN was designed for speed, not maximum sensitivity, and may not find distantly related coding sequences. BLASTX takes a nucleotide sequence, translates it in three forward reading frames and three reverse complement reading frames, and then compares the six translations against a protein sequence database. BLASTX is useful for sensitive analysis of preliminary (single-pass) sequence data and is tolerant of sequencing errors (Gish and States, *Nature Genetics, 3:* 266-272 (1993), the entirety of which is herein incorporated by reference). BLASTN and BLASTX may be used in concert for analyzing EST data (Coulson, *Trends in Biotechnology, 12:* 76-80 (1994); Birren, *et al., Genome Analysis, 1:* 543-559 (1997).

Given a coding nucleotide sequence and the protein it encodes, it is often preferable to use the protein as the query sequence to search a database because of the greatly increased sensitivity to detect more subtle relationships. This is due to the larger alphabet of proteins (20 amino acids) compared with the alphabet of nucleic acid sequences (4 bases), where it is far easier to obtain a match by chance. In addition, with nucleotide alignments, only a match

(positive score) or a mismatch (negative score) is obtained, but with proteins, the presence of conservative amino acid substitutions can be taken into account. Here, a mismatch may yield a positive score if the non-identical residue has physical/chemical properties similar to the one it replaced. Various scoring matrices are used to supply the substitution scores of all possible amino acid pairs. A general purpose scoring system is the BLOSUM62 matrix (Henikoff and Henikoff, *Proteins, 17:* 49-61 (1993), the entirety of which is herein incorporated by reference), which is currently the default choice for BLAST programs. BLOSUM62 is tailored for alignments of moderately diverged sequences and thus may not yield the best results under all conditions. Altschul, *J. Mol. Biol. 36:* 290-300 (1993), the entirety of which is herein incorporated by reference, uses a combination of three matrices to cover all contingencies. This may improve sensitivity, but at the expense of slower searches. In practice, a single BLOSUM62 matrix is often used but others (PAM40 and PAM250) may be attempted when additional analysis is necessary. Low PAM matrices are directed at detecting very strong but localized sequence similarities, whereas high PAM matrices are directed at detecting long but weak alignments between very distantly related sequences.

Homologues in other organisms are available that can be used for comparative sequence analysis. Multiple alignments are performed to study similarities and differences in a group of related sequences. CLUSTAL W is a multiple sequence alignment package available that performs progressive multiple sequence alignments based on the method of Feng and Doolittle, *J. Mol. Evol.* 25: 351-360 (1987), the entirety of which is herein incorporated by reference. Each pair of sequences is aligned and the distance between each pair is calculated; from this distance matrix, a guide tree is calculated, and all of the sequences are progressively aligned based on this tree. A feature of the program is its sensitivity to the effect of gaps on the alignment; gap penalties are varied to encourage the insertion of gaps in probable loop regions instead of in the middle of structured regions. Users can specify gap penalties, choose between a number of scoring matrices, or supply their own scoring matrix for both the pairwise alignments and the multiple alignments. CLUSTAL W for UNIX and VMS systems is available at: ftp.ebi.ac.uk.

Another program is MACAW (Schuler *et al.*, *Proteins, Struct. Func. Genet, 9:*180-190 (1991), the entirety of which is herein incorporated by reference, for which both Macintosh and Microsoft Windows versions are available. MACAW uses a graphical interface, provides a choice of several alignment algorithms, and is available by anonymous ftp at: ncbi.nlm.nih.gov (directory/pub/macaw).

Sequence motifs are derived from multiple alignments and can be used to examine individual sequences or an entire database for subtle patterns. With motifs, it is sometimes possible to detect distant relationships that may not be demonstrable based on comparisons of primary sequences alone. Currently, the largest collection of sequence motifs in the world is PROSITE (Bairoch and Bucher, *Nucleic Acid Research, 22:* 3583-3589 (1994), the entirety of which is herein incorporated by reference.) PROSITE may be accessed via either the ExPASy server on the World Wide Web or anonymous ftp site. Many commercial sequence analysis packages also provide search programs that use PROSITE data.

A resource for searching protein motifs is the BLOCKS E-mail server developed by S. Henikoff, *Trends Biochem Sci., 18:*267-268 (1993), the entirety of which is herein incorporated by reference; Henikoff and Henikoff, *Nucleic Acid Research, 19:*6565-6572 (1991), the entirety of which is herein incorporated by reference; Henikoff and Henikoff, *Proteins, 17:* 49-61 (1993). BLOCKS searches a protein or nucleotide sequence against a database of protein motifs or "blocks." Blocks are defined as short, ungapped multiple alignments that represent highly conserved protein patterns. The blocks themselves are derived from entries in PROSITE as well as other sources. Either a protein or nucleotide query can be submitted to the BLOCKS server; if a nucleotide sequence is submitted, the sequence is translated in all six reading frames and motifs are sought in these conceptual translations. Once the search is completed, the server will return a ranked list of significant matches, along with an alignment of the query sequence to the matched BLOCKS entries.

Conserved protein domains can be represented by two-dimensional matrices, which measure either the frequency or probability of the occurrences of each amino acid residue and

deletions or insertions in each position of the domain. This type of model, when used to search against protein databases, is sensitive and usually yields more accurate results than simple motif searches. Two popular implementations of this approach are profile searches (such as GCG program ProfileSearch) and Hidden Markov Models (HMMs)(Krough *et al., J. Mol. Biol.* *235*:1501-1531 (1994); Eddy, *Current Opinion in Structural Biology 6*:361-365 (1996), both of which are herein incorporated by reference in their entirety). In both cases, a large number of common protein domains have been converted into profiles, as present in the PROSITE library, or HHM models, as in the Pfam protein domain library (Sonnhammer *et al., Proteins 28*:405-420 (1997), the entirety of which is herein incorporated by reference). Pfam contains more than 500 HMM models for enzymes, transcription factors, signal transduction molecules, and structural proteins. Protein databases can be queried with these profiles or HMM models, which will identify proteins containing the domain of interest. For example, HMMSW or HMMFS, two programs in a public domain package called HMMER (Sonnhammer *et al., Proteins 28*:405-420 (1997)) can be used.

PROSITE and BLOCKS represent collected families of protein motifs. Thus, searching these databases entails submitting a single sequence to determine whether or not that sequence is similar to the members of an established family. Programs working in the opposite direction compare a collection of sequences with individual entries in the protein databases. An example of such a program is the Motif Search Tool, or MoST (Tatusov *et al. Proc. Natl. Acad. Sci. 91:* 12091-12095 (1994), the entirety of which is herein incorporated by reference.) On the basis of an aligned set of input sequences, a weight matrix is calculated by using one of four methods (selected by the user); a weight matrix is simply a representation, position by position in an alignment, of how likely a particular amino acid will appear. The calculated weight matrix is then used to search the databases. To increase sensitivity, newly found sequences are added to the original data set, the weight matrix is recalculated, and the search is performed again. This procedure continues until no new sequences are found.

Summary of the Invention

The present invention provides a substantially purified nucleic acid molecule that encodes a maize protein or fragment thereof comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 32236.

The present invention also provides one or more substantially purified nucleic acid molecules comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof.

The present invention also provides a substantially purified maize protein or fragment thereof, wherein said maize protein is encoded by a nucleic acid molecule that comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 32236.

The present invention further provides a substantially purified protein, peptide, or fragment thereof encoded by a nucleic acid sequence which specifically hybridizes to a nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO: 1 through SEQ ID NO:32236.

The present invention further provides a substantially purified antibody capable of specifically binding to a protein or fragment thereof encoded by a nucleic acid sequence which specifically hybridizes to a nucleic acid molecule having a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO:1 through SEQ ID NO:32236.

The present invention also provides a transformed plant transformed to contain a nucleic acid molecule which comprises: (A) an exogenous promoter region which functions in plant cells to cause the production of an mRNA molecule; which is linked to (B) a structural nucleic acid molecule, wherein said structural nucleic acid molecule comprises a nucleic acid molecule that encodes a protein, peptide, or fragment thereof which hybridizes to a nucleic acid sequence selected from the group consisting of a complement of SEQ ID NO:1 through SEQ ID NO:32236 expressed in an effective amount to produce a desirable agronomic effect; which is linked to (C)

a 3' non-translated sequence that functions in plant cells to cause the termination of transcription and the addition of polyadenylated ribonucleotides to the 3' end of the mRNA sequence.

The present invention also provides a transformed plant cell containing a nucleic acid molecule whose non-transcribed strand encodes a protein or fragment thereof, wherein the transcribed strand of said nucleic acid is complementary to a nucleic acid molecule that encodes a protein or fragment thereof. The present invention also provides bacterial, viral, microbial, and plant cells comprising a nucleic acid molecule of the present invention

The present invention also provides a method of producing a plant containing one or more proteins encoded by sequences comprising SEQ ID NO:1 or complement thereof through SEQ ID NO:32236 or complements thereof, expressed in a sufficient amount and/or fashion to produce a desirable agronomic effect.

In accomplishing the foregoing, there is provided, in accordance with one aspect of the present invention, methods of producing genetically transformed plants, comprising the steps of:

(a) inserting into the genome of a plant cell a recombinant, double-stranded DNA molecule comprising

(I) a promoter which functions in plant cells to cause the production of an RNA sequence,

(ii) a structural DNA sequence that causes the production of an RNA sequence which encodes a desired protein.

(iii) a 3' non-translated DNA sequence which functions in plant cells to cause the addition of polyadenylated nucleotides to the 3' end of RNA sequence; where the promoter is homologous or heterologous with respect to the coding sequence and adapted to cause sufficient expression of a protein in desired plant tissues to enhance the agronomic utility of a plant transformed with said gene.

(b) obtaining a transformed plant cell with said nucleic acid molecule that encodes one or more proteins, wherein said nucleic acid molecule is transcribed and results in expression of said protein(s); and

(c) regenerating from the transformed plant cell a genetically transformed plant

The present invention also encompasses differentiated plants, seeds, and progeny comprising said transformed plant cells and which exhibit novel properties of agronomic significance.

The present invention also provides a method of producing a plant containing reduced levels of a protein comprising: (A) transforming a plant cell with a nucleic acid molecule that encodes a protein, wherein said nucleic acid molecule is transcribed and results in co-suppression of endogenous protein synthesis activity, and (B) regenerating plants and producing subsequent progeny from the transformed plant.

The present invention also provides a method of determining an association between a polymorphism and a plant trait comprising: (A) hybridizing a nucleic acid molecule specific for a polymorphism to genetic material of a plant, wherein said nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof; and (B) calculating the degree of association between the polymorphism and the plant trait.

The present invention also provides a method of isolating a genetic region, or nucleic acid that encodes a protein or fragment thereof comprising: (A) incubating under conditions permitting nucleic acid hybridization: a marker nucleic acid molecule, preferably an EST, with a complementary nucleic acid molecule obtained from a plant cell or plant tissue; (B) permitting hybridization between said marker nucleic acid molecule, preferably an EST, and said complementary nucleic acid molecule obtained from said plant cell or plant tissue; and (C) isolating said complementary nucleic acid molecule.

The present invention also provides a method for determining a level or pattern in a plant cell of a protein in a plant comprising: (A) incubating, under conditions permitting nucleic acid

hybridization, a marker nucleic acid molecule, the marker nucleic acid molecule selected from the group of marker nucleic acid molecules which specifically hybridize to a nucleic acid molecule having the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 32236 or complements thereof or fragments of either, with a complementary nucleic acid molecule obtained from the plant cell or plant tissue, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant tissue permits the detection of an mRNA for the enzyme; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant tissue; and (C) detecting the level or pattern of the complementary nucleic acid, wherein the detection of the complementary nucleic acid is predictive of the level or pattern of the protein.

The present invention also provides a method for determining the level or pattern of a protein in a plant cell or plant tissue comprising: (A) incubating under conditions permitting nucleic acid hybridization: a marker nucleic acid molecule, the marker nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof, with a complementary nucleic acid molecule obtained from a plant cell or plant tissue, wherein nucleic acid hybridization between the marker nucleic acid molecule, and the complementary nucleic acid molecule obtained from the plant cell or plant tissue permits the detection of said protein; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant cell or plant tissue; and (C) detecting the level or pattern of the complementary nucleic acid, wherein the detection of said complementary nucleic acid is predictive of the level or pattern of the protein synthesis.

The present invention also provides a method for determining a level or pattern of a protein in a plant cell or plant tissue which comprises assaying the concentration of a molecule, whose concentration is dependent upon the expression of a gene, the gene having a nucleic acid sequence which specifically hybridizes to a protein marker nucleic acid molecule, the molecule

being present in a plant cell or plant tissue, in comparison to the concentration of that molecule present in a plant cell or plant tissue with a known level or pattern of said protein, wherein an assayed concentration of the molecule is compared to the assayed concentration of the molecule in a plant cell or plant tissue with a known level or pattern of said protein.

The present invention also provides a method of determining a mutation in a plant whose presence is predictive of a mutation affecting a level or pattern of a protein comprising the steps: (A) incubating, under conditions permitting nucleic acid hybridization, a marker nucleic acid, the marker nucleic acid selected from the group of marker nucleic acid molecules which specifically hybridize to a nucleic acid molecule consisting of the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 32236 or complements thereof or fragments of either and a complementary nucleic acid molecule obtained from the plant, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant permits the detection of a polymorphism whose presence is predictive of a mutation affecting the level or pattern of the protein in the plant; (B) permitting hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant; and (C) detecting the presence of the polymorphism, wherein the detection of the polymorphism is predictive of the mutation.

The present invention also provides a method for determining a mutation in a plant whose presence is predictive of a mutation affecting the level or pattern of protein synthesis comprising the steps: (A) incubating under conditions permitting nucleic acid hybridization: a marker nucleic acid molecule, the marker nucleic acid molecule comprising a nucleic acid molecule that is linked to gene, the gene having a nucleic acid sequence which specifically hybridizes to a sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO:32236 and complements thereof, and a complementary nucleic acid molecule obtained from a plant tissue or plant cell of the plant, wherein nucleic acid hybridization between the marker nucleic acid molecule and the complementary nucleic acid molecule obtained from the plant permits the detection of a polymorphism whose presence is predictive of a mutation affecting said level or

pattern of a protein synthesis in the plant; (B) permitting hybridization between said marker nucleic acid molecule and said complementary nucleic acid molecule obtained from said plant; and; (C) detecting the presence of the polymorphism, wherein the detection of the polymorphism is predictive of the mutation.

The present invention also provides a method for reducing expression of a protein in a plant cell, the method comprising: growing a transformed plant cell containing a nucleic acid molecule whose non-transcribed strand encodes a protein or fragment thereof, wherein the transcribed strand of said nucleic acid is complementary to a nucleic acid molecule that encodes the protein in said plant cell, and whereby the strand that is complementary to the nucleic acid molecule that encodes the protein reduces or depresses expression of the protein.

The present invention provides maize nucleic acid molecules for use as molecular tags to isolate genetic regions (*i.e.* promoters and flanking sequences), isolate genes, map genes, and determine gene function. The present invention further provides maize nucleic acid molecules for use in determining if genes are members of a particular gene family.

The present invention also provides a method of obtaining full length genes using maize ESTs or complements thereof or fragments of either.

The present invention also provides a method of isolating promoters and flanking sequences using maize ESTs or complements thereof or fragments of either.

The present invention also provides maize ESTs or complements thereof or fragments of either for use in marker-assisted breeding programs.

The present invention also provides a method of identifying tissues comprising hybridizing nucleic acids from the tissue with maize ESTs or complements thereof or fragments of either.

The present invention also provides a method for production of antibodies targeted against the proteins, peptides, or fragments produced by the disclosed or complements thereof or fragments of either.

The present invention also provides a method for the transformation and regeneration of plants comprising sequences hybridizable to the disclosed ESTs or complements thereof or fragments of either.

The present invention also provides a method of modifying plant protein expression by inserting in a chimeric gene sense or antisense constructs of the maize ESTs.

Detailed Description of the Invention

Agents

(a) Nucleic Acid Molecules

Agents of the present invention include nucleic acid molecules and more specifically EST nucleic acid molecules or nucleic acid fragment molecules thereof. Fragment EST nucleic acid molecules may encode significant portion(s) of, or indeed most of, the EST nucleic acid molecule. Alternatively, the fragments may comprise smaller oligonucleotides (having from about 15 to about 250 nucleotide residues, and more preferably, about 15 to about 30 nucleotide residues).

A subset of the nucleic acid molecules of the present invention includes nucleic acid molecules that are marker molecules. Another subset of the nucleic acid molecules of the present invention include nucleic acid molecules that encode a protein or fragment thereof. Another subset of the nucleic acid molecules of the present invention are EST molecules.

The term "substantially purified", as used herein, refers to a molecule separated from substantially all other molecules normally associated with it in its native state. More preferably a substantially purified molecule is the predominant species present in a preparation. A substantially purified molecule may be greater than 60% free, preferably 75% free, more preferably 90% free, and most preferably 95% free from the other molecules (exclusive of solvent) present in the natural mixture. The term "substantially purified" is not intended to encompass molecules present in their native state.

The agents of the present invention will preferably be "biologically active" with respect to either a structural attribute, such as the capacity of a nucleic acid to hybridize to another nucleic acid molecule, or the ability of a protein to be bound by antibody (or to compete with another molecule for such binding). Alternatively, such an attribute may be catalytic, and thus involve the capacity of the agent to mediate a chemical reaction or response.

The agents of the present invention may also be recombinant. As used herein, the term recombinant means any agent (*e.g.* DNA, peptide *etc.*), that is, or results, however indirect, from human manipulation of a nucleic acid molecule.

It is understood that the agents of the present invention may be labeled with reagents that facilitate detection of the agent (*e.g.* fluorescent labels (Prober, *et al.*, *Science* 238:336-340 (1987); Albarella *et al.*, EP 144914, chemical labels (Sheldon *et al.*, U.S. Patent 4,582,789; Albarella *et al.*, U.S. Patent 4,563,417, modified bases (Miyoshi *et al.*, EP 119448, all of which are hereby incorporated by reference in their entirety).

It is further understood, that the present invention provides bacterial, viral, microbial, and plant cells comprising the agents of the present invention.

Nucleic acid molecules or fragment thereof of the present invention are capable of specifically hybridizing to other nucleic acid molecules under certain circumstances. As used herein, two nucleic acid molecules are said to be capable of specifically hybridizing to one another if the two molecules are capable of forming an anti-parallel, double-stranded nucleic acid structure. A nucleic acid molecule is said to be the "complement" of another nucleic acid molecule if they exhibit complete complementarity. As used herein, molecules are said to exhibit "complete complementarity" when every nucleotide of one of the molecules is complementary to a nucleotide of the other. Two molecules are said to be "minimally complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under at least conventional "low-stringency" conditions. Similarly, the molecules are said to be "complementary" if they can hybridize to one another with sufficient stability to permit them to remain annealed to one another under conventional "high-

stringency" conditions. Conventional stringency conditions are described by Sambrook, *et al.*, In: *Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press*, Cold Spring Harbor, New York (1989), and by Haymes, *et al.* In: *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985), the entirety of which is herein incorporated by reference. Departures from complete complementarity are therefore permissible, as long as such departures do not completely preclude the capacity of the molecules to form a double-stranded structure. Thus, in order for an nucleic acid molecule or fragment of the present invention to serve as a primer or probe it need only be sufficiently complementary in sequence to be able to form a stable double-stranded structure under the particular solvent and salt concentrations employed.

Appropriate stringency conditions which promote DNA hybridization are, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C. Both temperature and salt may be varied, or either the temperature or the salt concentration may be held constant while the other variable is changed.

In a preferred embodiment, a nucleic acid of the present invention will specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO: 1 through SEQ ID NO: 32236 or complements thereof under moderately stringent conditions, for example, at about 2.0 x SSC and about 65°C.

In a particularly preferred embodiment, a nucleic acid of the present invention will include those nucleic acid molecules that specifically hybridize to one or more of the nucleic acid molecules set forth in SEQ ID NO:1 through SEQ ID NO: 32236 or complements thereof under high stringency conditions.

In one aspect of the present invention, the nucleic acid molecules of the present invention have one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof. In another aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 90% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof. In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 95% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof. In a more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 98% sequence identity with one or more of the nucleic acid sequences set forth in SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof. In an even more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention share between 100% and 99% sequence identity with one or more of the sequences set forth in SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof. In a further, even more preferred aspect of the present invention, one or more of the nucleic acid molecules of the present invention exhibit 100% sequence identity with one or more nucleic acid molecules present within the cDNA libraries LIB3115, LIB3116, LIB3117, LIB3118, LIB3150, LIB3151, LIB3152, LIB3153, LIB3154, LIB3180, LIB3181, and LIB3182 (Monsanto Company, St. Louis, Missouri, United States of America).

In a preferred embodiment of the present invention, a maize protein or fragment thereof of the present invention is a homologue of another plant protein. In another preferred embodiment of the present invention, a maize protein or fragment thereof of the present invention is a homologue of a fungal protein. In another preferred embodiment of the present invention, a maize protein or fragment thereof of the present invention is a homologue of a mammalian protein. In another preferred embodiment of the present invention, a maize protein or fragment thereof of the present invention is a homologue of an algal protein. In another preferred

embodiment of the present invention, a maize protein or fragment thereof of the present invention is a homologue of a bacterial protein. In another preferred embodiment of the present invention, a maize protein or fragment thereof of the present invention is a homologue of a soybean protein.

In a preferred embodiment of the present invention, the nucleic molecule of the present invention encodes a maize protein or fragment thereof where a maize protein or fragment thereof exhibits a BLAST probability score of greater than 1E-12, preferably a BLAST probability score of between about 1E-30 and about 1E-12, even more preferably a BLAST probability score of greater than 1E-30 with its homologue.

In another preferred embodiment of the present invention, the nucleic acid molecule encoding a maize protein or fragment thereof exhibits a % identity with its homologue of between about 25% and about 40%, more preferably of between about 40% and about 70%, even more preferably of between about 70% and about 90% and even more preferably between about 90% and 99%. In another preferred embodiment, of the present invention, a maize protein or fragment thereof exhibits a % identity with its homologue of 100%.

In a preferred embodiment of the present invention, the nucleic acid molecule of the present invention encodes a maize protein or fragment thereof where the maize protein exhibits a BLAST score of greater than 120, preferably a BLAST score of between about 1450 and about 120, even more preferably a BLAST score of greater than 1450 with its homologue.

Nucleic acid molecules of the present invention also include non-maize homologues. Preferred non-maize homologues are selected from the group consisting of alfalfa, *Arabidopsis*, barley, *Brassica*, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, an ornamental plant, pea, peanut, pepper, potato, rice, rye, sorghum, soybean strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, lentils, grape, banana, tea, turf grasses, sunflower, oil palm and *Phaseolus*.

The degeneracy of the genetic code, which allows different nucleic acid sequences to code for the same protein or peptide, is known in the literature. (U.S. Patent No. 4,757,006, the entirety of which is herein incorporated by reference).

In an aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a maize protein or fragment thereof in SEQ ID NO: 1 through SEQ ID NO: 32236 due to the degeneracy in the genetic code in that they encode the same protein but differ in nucleic acid sequence.

In another further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a maize protein or fragment thereof in SEQ ID NO: 1 through SEQ ID NO: 32236 due to fact that the different nucleic acid sequences encode a protein having one or more conservative amino acid residues. It is understood that codons capable of coding for such conservative substitutions are known in the art.

It is well known in the art that one or more amino acids in a native sequence can be substituted with another amino acid(s), the charge and polarity of which are similar to that of the native amino acid, *i.e.*, a conservative amino acid substitution, resulting in a silent change. Conserved substitutes for an amino acid within the native polypeptide sequence can be selected from other members of the class to which the naturally occurring amino acid belongs. Amino acids can be divided into the following four groups: (1) acidic amino acids, (2) basic amino acids, (3) neutral polar amino acids, and (4) neutral nonpolar amino acids. Representative amino acids within these various groups include, but are not limited to, (1) acidic (negatively charged) amino acids such as aspartic acid and glutamic acid; (2) basic (positively charged) amino acids such as arginine, histidine, and lysine; (3) neutral polar amino acids such as glycine, serine, threonine, cysteine, cystine, tyrosine, asparagine, and glutamine; and (4) neutral nonpolar (hydrophobic) amino acids such as alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine.

Conservative amino acid changes within the native polypeptides sequence can be made by substituting one amino acid within one of these groups with another amino acid within the same group. Biologically functional equivalents of the proteins or fragments thereof of the present invention can have 10 or fewer conservative amino acid changes, more preferably seven or fewer conservative amino acid changes, and most preferably five or fewer conservative amino acid changes. The encoding nucleotide sequence will thus have corresponding base substitutions, permitting it to encode biologically functional equivalent forms of the proteins or fragments of the present invention.

It is understood that certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Because it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence substitutions can be made in a protein sequence and, of course, its underlying DNA coding sequence and, nevertheless, obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the proteins or fragments of the present invention, or corresponding DNA sequences that encode said peptides, without appreciable loss of their biological utility or activity. It is understood that codons capable of coding for such amino acid changes are known in the art.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte and Doolittle, *J. Mol. Biol. 157*, 105-132 (1982), herein incorporated by reference in its entirety). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982); these are isoleucine (+4.5), valine (+4.2), leucine (+3.8), phenylalanine (+2.8), cysteine/cystine (+2.5), methionine (+1.9), alanine (+1.8), glycine (-0.4), threonine (-0.7), serine (-0.8), tryptophan (-0.9), tyrosine (-1.3), proline (-1.6), histidine (-3.2), glutamate (-3.5), glutamine (-3.5), aspartate (-3.5), asparagine (-3.5), lysine (-3.9), and arginine (-4.5).

In making such changes, the substitution of amino acids whose hydropathic indices are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent 4,554,101, incorporated herein by reference in its entirety, states that the greatest local average hydrophilicity of a protein, as govern by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U.S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0), lysine (+3.0), aspartate (+3.0±1), glutamate (+3.0±1), serine (+0.3), asparagine (+0.2), glutamine (+0.2), glycine (0), threonine (-0.4), proline (-0.5±1), alanine (-0.5), histidine (-0.5), cysteine (-1.0), methionine (-1.3), valine (-1.5), leucine (-1.8), isoleucine (-1.8), tyrosine (-2.3), phenylalanine (-2.5), and tryptophan (-3.4).

In making such changes, the substitution of amino acids whose hydrophilicity values are within ±2 is preferred, those which are within ±1 are particularly preferred, and those within ±0.5 are even more particularly preferred. In a further aspect of the present invention, one or more of the nucleic acid molecules of the present invention differ in nucleic acid sequence from those encoding a maize protein or fragment thereof set forth in SEQ ID NO: 1 through SEQ ID NO: 32236 or fragment thereof due to the fact that one or more codons encoding an amino acid has been substituted for a codon that encodes a nonessential substitution of the amino acid originally encoded.

One aspect of the present invention concerns markers that include nucleic acid molecules SEQ ID NO: 1 through SEQ ID NO: 32236 or complements thereof or fragments of either that can act as markers or other nucleic acid molecules of the present invention that can act as markers. Genetic markers of the present invention include "dominant" or "codominant" markers "Codominant markers" reveal the presence of two or more alleles (two per diploid individual) at a locus. "Dominant markers" reveal the presence of only a single allele per locus. The presence of the dominant marker phenotype (*e.g.*, a band of DNA) is an indication that one allele is present in either the homozygous or heterozygous condition. The absence of the dominant marker phenotype (*e.g.* absence of a DNA band) is merely evidence that "some other" undefined allele is present. In the case of populations where individuals are predominantly homozygous and loci are predominately dimorphic, dominant and codominant markers can be equally valuable. As populations become more heterozygous and multi-allelic, codominant markers often become more informative of the genotype than dominant markers. Marker molecules can be, for example, capable of detecting polymorphisms such as single nucleotide polymorphisms (SNPs).

SNPs are single base changes in genomic DNA sequence. They occur at greater frequency and are spaced with a greater uniformity throughout a genome than other reported forms of polymorphism. The greater frequency and uniformity of SNPs means that there is greater probability that such a polymorphism will be found near or in a genetic locus of interest than would be the case for other polymorphisms. SNPs are located in protein-coding regions and noncoding regions of a genome. Some of these SNPs may result in defective or variant protein expression (*e.g.*, as a results of mutations or defective splicing). Analysis (genotyping) of characterized SNPs can require only a plus/minus assay rather than a lengthy measurement, permitting easier automation.

SNPs can be characterized using any of a variety of methods. Such methods include the direct or indirect sequencing of the site, the use of restriction enzymes (Botstein *et al.*, *Am. J. Hum. Genet. 32*:314-331 (1980), the entirety of which is herein incorporated reference;

Konieczny and Ausubel, *Plant J. 4*:403-410 (1993), the entirety of which is herein incorporated by reference), enzymatic and chemical mismatch assays (Myers *et al.*, *Nature 313*:495-498 (1985), the entirety of which is herein incorporated by reference), allele-specific PCR (Newton *et al.*, *Nucl. Acids Res. 17*:2503-2516 (1989), the entirety of which is herein incorporated by reference; Wu *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 86*:2757-2760 (1989), the entirety of which is herein incorporated by reference), ligase chain reaction (Barany, *Proc. Natl. Acad. Sci. (U.S.A.) 88*:189-193 (1991), the entirety of which is herein incorporated by reference), single-strand conformation polymorphism analysis (Labrune *et al.*, *Am. J. Hum. Genet. 48*: 1115-1120 (1991), the entirety of which is herein incorporated by reference), primer-directed nucleotide incorporation assays (Kuppuswami *et al.*, *Proc. Natl. Acad. Sci. USA 88*:1143-1147 (1991), the entirety of which is herein incorporated by reference), dideoxy fingerprinting (Sarkar *et al.*, *Genomics 13*:441-443 (1992), the entirety of which is herein incorporated by reference), solid-phase ELISA-based oligonucleotide ligation assays (Nikiforov *et al.*, *Nucl. Acids Res. 22*:4167-4175 (1994), the entirety of which is herein incorporated by reference), oligonucleotide fluorescence-quenching assays (Livak *et al.*, *PCR Methods Appl. 4*:357-362 (1995), the entirety of which is herein incorporated by reference), 5'-nuclease allele-specific hybridization TaqMan assay (Livak *et al.*, *Nature Genet. 9*:341-342 (1995), the entirety of which is herein incorporated by reference), template-directed dye-terminator incorporation (TDI) assay (Chen and Kwok, *Nucl. Acids Res. 25*:347-353 (1997), the entirety of which is herein incorporated by reference), allele-specific molecular beacon assay (Tyagi *et al.*, *Nature Biotech. 16*: 49-53 (1998), the entirety of which is herein incorporated by reference), PinPoint assay (Haff and Smirnov, *Genome Res. 7*: 378-388 (1997), the entirety of which is herein incorporated by reference) and dCAPS analysis (Neff *et al.*, *Plant J. 14*:387-392 (1998), the entirety of which is herein incorporated by reference).

Additional markers, such as AFLP markers, RFLP markers and RAPD markers, can be utilized (Walton, *Seed World* 22-29 (July, 1993), the entirety of which is herein incorporated by reference; Burow and Blake, *Molecular Dissection of Complex Traits*, 13-29, Paterson (ed.),

CRC Press, New York (1988), the entirety of which is herein incorporated by reference). DNA markers can be developed from nucleic acid molecules using restriction endonucleases, the PCR and/or DNA sequence information. RFLP markers result from single base changes or insertions/deletions. These codominant markers are highly abundant in plant genomes, have a medium level of polymorphism and are developed by a combination of restriction endonuclease digestion and Southern blotting hybridization. CAPS are similarly developed from restriction nuclease digestion but only of specific PCR products. These markers are also codominant, have a medium level of polymorphism and are highly abundant in the genome. The CAPS result from single base changes and insertions/deletions.

Another marker type, RAPDs, are developed from DNA amplification with random primers and result from single base changes and insertions/deletions in plant genomes. They are dominant markers with a medium level of polymorphisms and are highly abundant. AFLP markers require using the PCR on a subset of restriction fragments from extended adapter primers. These markers are both dominant and codominant are highly abundant in genomes and exhibit a medium level of polymorphism.

SSRs require DNA sequence information. These codominant markers result from repeat length changes, are highly polymorphic and do not exhibit as high a degree of abundance in the genome as CAPS, AFLPs and RAPDs, SNPs also require DNA sequence information. These codominant markers result from single base substitutions. They are highly abundant and exhibit a medium of polymorphism (Rafalski *et al.*, In: *Nonmammalian Genomic Analysis*, Birren and Lai (ed.), Academic Press, San Diego, CA, pp. 75-134 (1996), the entirety of which is herein incorporated by reference). It is understood that a nucleic acid molecule of the present invention may be used as a marker.

A PCR probe is a nucleic acid molecule capable of initiating a polymerase activity while in a double-stranded structure with another nucleic acid. Various methods for determining the structure of PCR probes and PCR techniques exist in the art. Computer generated searches using programs such as Primer3 (www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi), STSPipeline

(www-genome.wi.mit.edu/cgi-bin/www-STS_Pipeline), or GeneUp (Pesole *et al.*, *BioTechniques* *25*:112-123 (1998) the entirety of which is herein incorporated by reference), for example, can be used to identify potential PCR primers.

It is understood that a fragment of one or more of the nucleic acid molecules of the present invention may be a probe and specifically a PCR probe.

(b) Protein and Peptide Molecules

A class of agents comprises one or more of the protein or peptide molecules encoded by SEQ ID NO: 1 through SEQ ID NO:32236 or one or more of the protein or fragment thereof or peptide molecules encoded by other nucleic acid agents of the present invention. As used herein, the term "protein molecule" or "peptide molecule" includes any molecule that comprises five or more amino acids. It is well know in the art that proteins may undergo modification, including post-translational modifications, such as, but not limited to, disulfide bond formation, glycosylation, phosphorylation, or oligomerization. Thus, as used herein, the term "protein molecule" or "peptide molecule" includes any protein molecule that is modified by any biological or non-biological process. The terms "amino acid" and "amino acids" refer to all naturally occurring L-amino acids. This definition is meant to include norleucine, ornithine, homocysteine, and homoserine.

One or more of the protein or fragment of peptide molecules may be produced via chemical synthesis, or more preferably, by expression in a suitable bacterial or eukaryotic host. Suitable methods for expression are described by Sambrook, *et al.*, (In: *Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press*, Cold Spring Harbor, New York (1989)), or similar texts.

A "protein fragment" is a peptide or polypeptide molecule whose amino acid sequence comprises a subset of the amino acid sequence of that protein. A protein or fragment thereof that comprises one or more additional peptide regions not derived from that protein is a "fusion" protein. Such molecules may be derivatized to contain carbohydrate or other moieties (such as

keyhole limpet hemocyanin, *etc.*). Fusion protein or peptide molecule of the present invention are preferably produced via recombinant means.

Another class of agents comprise protein or peptide molecules encoded by SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof or, fragments or fusions thereof in which non-essential, or not relevant, amino acid residues have been added, replaced, or deleted. An example of such a homologue is the homologue protein of all non-maize plant species, including but not limited to alfalfa, *Arabidopsis*, barley, *Brassica*, broccoli, cabbage, citrus, cotton, garlic, oat, oilseed rape, onion, canola, flax, an ornamental plant, pea, peanut, pepper, potato, rice, rye, sorghum, soybean, strawberry, sugarcane, sugarbeet, tomato, wheat, poplar, pine, fir, eucalyptus, apple, lettuce, peas, lentils, grape, banana, tea, turf grasses, *etc.* Particularly preferred non-maize plants to utilize for the isolation of homologues would include alfalfa, *Arabidopsis*, barley, cotton, oat, oilseed rape, rice, canola, ornamentals, soybean, sugarcane, sugarbeet, tomato, potato, wheat, and turf grasses. Such a homologue can be obtained by any of a variety of methods. Most preferably, as indicated above, one or more of the disclosed sequences (SEQ ID NO: 1 through SEQ ID NO:32236 or complements thereof) will be used to define a pair of primers that may be used to isolate the homologue-encoding nucleic acid molecules from any desired species. Such molecules can be expressed to yield homologues by recombinant means.

(c) Antibodies

One aspect of the present invention concerns antibodies, single-chain antigen binding molecules, or other proteins that specifically bind to one or more of the protein or peptide molecules of the present invention and their homologues, fusions or fragments. Such antibodies may be used to quantitatively or qualitatively detect the protein or peptide molecules of the present invention. As used herein, an antibody or peptide is said to "specifically bind" to a protein or peptide molecule of the present invention if such binding is not competitively inhibited by the presence of non-related molecules.

Nucleic acid molecules that encode all or part of the protein of the present invention can be expressed, via recombinant means, to yield protein or peptides that can in turn be used to elicit

antibodies that are capable of binding the expressed protein or peptide. Such antibodies may be used in immunoassays for that protein. Such protein-encoding molecules, or their fragments may be a "fusion" molecule (*i.e.*, a part of a larger nucleic acid molecule) such that, upon expression, a fusion protein is produced. It is understood that any of the nucleic acid molecules of the present invention may be expressed, via recombinant means, to yield proteins or peptides encoded by these nucleic acid molecules.

The antibodies that specifically bind proteins and protein fragments of the present invention may be polyclonal or monoclonal, and may comprise intact immunoglobulins, or antigen binding portions of immunoglobulins (such as (F(ab'), $F(ab')_2$) fragments, or single-chain immunoglobulins producible, for example, via recombinant means). It is understood that practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of antibodies (see, for example, Harlow and Lane, In *Antibodies: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, New York (1988), the entirety of which is herein incorporated by reference).

Murine monoclonal antibodies are particularly preferred. BALB/c mice are preferred for this purpose, however, equivalent strains may also be used. The animals are preferably immunized with approximately 25 μg of purified protein (or fragment thereof) that has been emulsified a suitable adjuvant (such as TiterMax adjuvant (Vaxcel, Norcross, GA)). Immunization is preferably conducted at two intramuscular sites, one intraperitoneal site, and one subcutaneous site at the base of the tail. An additional i.v. injection of approximately 25 μg of antigen is preferably given in normal saline three weeks later. After approximately 11 days following the second injection, the mice may be bled and the blood screened for the presence of anti-protein or peptide antibodies. Preferably, a direct binding Enzyme-Linked Immunoassay (ELISA) is employed for this purpose.

More preferably, the mouse having the highest antibody titer is given a third i.v. injection of approximately 25 μg of the same protein or fragment. The splenic leukocytes from this animal may be recovered 3 days later, and are then permitted to fuse, most preferably, using

polyethylene glycol, with cells of a suitable myeloma cell line (such as, for example, the P3X63Ag8.653 myeloma cell line). Hybridoma cells are selected by culturing the cells under "HAT" (hypoxanthine-aminopterin-thymine) selection for about one week. The resulting clones may then be screened for their capacity to produce monoclonal antibodies ("mAbs), preferably by direct ELISA.

In one embodiment, anti-protein or peptide monoclonal antibodies are isolated using a fusion of a protein, protein fragment, or peptide of the present invention, or conjugate of a protein, protein fragment, or peptide of the present invention, as immunogens. Thus, for example, a group of mice can be immunized using a fusion protein emulsified in Freund's complete adjuvant (*e.g.* approximately 50 μg of antigen per immunization). At three week intervals, an identical amount of antigen is emulsified in Freund's incomplete adjuvant and used to immunize the animals. Ten days following the third immunization, serum samples are taken and evaluated for the presence of antibody. If antibody titers are too low, a fourth booster can be employed. Polysera capable of binding the protein or peptide can also be obtained using this method.

In a preferred procedure for obtaining monoclonal antibodies, the spleens of the above-described immunized mice are removed, disrupted, and immune splenocytes are isolated over a ficoll gradient. The isolated splenocytes are fused, using polyethylene glycol with BALB/c-derived HGPRT (hypoxanthine guanine phosphoribosyl transferase) deficient P3x63xAg8.653 plasmacytoma cells. The fused cells are plated into 96-well microtiter plates and screened for hybridoma fusion cells by their capacity to grow in culture medium supplemented with hypothanthine, aminopterin and thymidine for approximately 2-3 weeks.

Hybridoma cells that arise from such incubation are preferably screened for their capacity to produce an immunoglobulin that binds to a protein of interest. An indirect ELISA may be used for this purpose. In brief, the supernatants of hybridomas are incubated in microtiter wells that contain immobilized protein. After washing, the titer of bound immunoglobulin can be determined using, for example, a goat anti-mouse antibody conjugated to horseradish peroxidase.

After additional washing, the amount of immobilized enzyme is determined (for example through the use of a chromogenic substrate). Such screening is performed as quickly as possible after the identification of the hybridoma in order to ensure that a desired clone is not overgrown by non-secreting neighbors. Desirably, the fusion plates are screened several times since the rates of hybridoma growth vary. In a preferred sub-embodiment, a different antigenic form of immunogen may be used to screen the hybridoma. Thus, for example, the splenocytes may be immunized with one immunogen, but the resulting hybridomas can be screened using a different immunogen. It is understood that any of the protein or peptide molecules of the present invention may be used to raise antibodies.

As discussed below, such antibody molecules or their fragments may be used for diagnostic purposes. Where the antibodies are intended for diagnostic purposes, it may be desirable to derivatize them, for example with a ligand group (such as biotin) or a detectable marker group (such as a fluorescent group, a radioisotope or an enzyme).

The ability to produce antibodies that bind the protein or peptide molecules of the present invention permits the identification of mimetic compounds of those molecules. A "mimetic compound" is a compound that is not that compound, or a fragment of that compound, but which nonetheless exhibits an ability to specifically bind to antibodies directed against that compound.

It is understood that any of the agents of the present invention can be substantially purified and/or be biologically active and/or recombinant.

Uses of the Agents of the Invention

The nucleic acid molecules and fragments thereof of the present invention from the normalized cDNA libraries LIB3115 and LIB3116 are isolated from maize (RX601, Asgrow Seed Company, Des Moines, Iowa U.S.A.) pooled leaf tissue harvested from field grown plants at Asgrow research stations. Leaves are the carbohydrate factories of crop plants, therefore, the ESTs of the present invention will find great use in the isolation of a variety of agronomically significant genes, including but not limited to genes that are necessary for the interception and transformation of light energy via photosynthesis linked with plant growth, quality and yield.

Genes isolated using the disclosed ESTs would also be involved in pathways, including but not limited to, of light and dark respiration, of CO_2 assimilation, and of nitrogen metabolism linked to fruiting and mobilization and distribution of nitrogen. The ESTs of the present invention also can enable the acquisition of molecular markers, which can be used in, including but not limited to, breeding schemes, genetic and molecular mapping, and cloning of agronomically significant genes.

The nucleic acid molecules and fragments thereof of the present invention from the normalized cDNA libraries LIB3117 and LIB3118 are isolated from maize (RX601, Asgrow Seed Company, Des Moines, Iowa U.S.A) pooled kernels from plants at 15 to 20 days after pollination. Libraries from this tissue can enable the acquisition of a variety of agronomically significant genes expressed at this kernel development stage. The ESTs of the present invention can enable the acquisition of, but are not limited to genes that regulate protein, early kernel development, cell division, amyloplast biogenesis, starch biosynthesis, sucrose biosynthesis, carbon transport, oils, amino acids, sterols, minerals, isoflavones, saponins, vitamins, tocopherols, antinutrient components, carbohydrates, kernel cytokinin metabolism and seed regulatory elements. Such genes are associated with plant growth, quality and yield, and could also serve as links in important developmental, metabolic, and catabolic pathways. The ESTs of the present invention also can enable the acquisition of kernel specific promoters and cis-regulatory elements which will be useful to express agronomically significant genes in these tissues and/or other tissues. The ESTs of the present invention also can enable the acquisition of molecular markers, which can be used in, including but not limited to, breeding schemes, genetic and molecular mapping, and cloning of agronomically significant genes.

The nucleic acid molecules and fragments thereof of the present invention from the cDNA libraries LIB3150, LIB3151, LIB3152, LIB3153, LIB3154, LIB3180, LIB3181, LIB3182 are isolated from *Zea mays* endosperm tissue. Libraries from this tissue can enable the acquisition of a variety of agronomically significant genes involved in endosperm carbohydrate metabolism and highly expressed during early maize endosperm development. The ESTs of the

present invention can enable the acquisition of, including but not limited to genes that regulate protein, oils, amino acids, sterols, minerals, isoflavones, saponins, vitamins, tocopherols, antinutrient components, carbohydrates, starch metabolism and seed regulatory elements. Such genes are associated with plant growth, quality and yield, and could also serve as links in important developmental, metabolic, and catabolic pathways. The ESTs of the present invention also can enable the acquisition of endosperm-specific promoters and cis-regulatory elements which will be useful to express agronomically significant genes in these tissues and/or other tissues. The ESTs of the present invention also can enable the acquisition of molecular markers, which can be used in, including but not limited to, breeding schemes, genetic and molecular mapping, and cloning of agronomically significant genes.

Nucleic acid molecules and fragments thereof of the present invention may be employed to obtain other nucleic acid molecules. Such molecules include the nucleic acid molecules of other plants or other organisms (*e.g.*, alfalfa, rice, potato, cotton, oat, rye, barley, maize, wheat, *Arabidopsis, Brassica, etc.*) including the nucleic acid molecules that encode, in whole or in part, protein homologues of other plant species or other organisms, and sequences of genetic elements such as promoters and transcriptional regulatory elements. Such molecules can be readily obtained by using the above-described nucleic acid molecules or fragments thereof to screen cDNA or genomic libraries obtained from such plant species. Methods for forming such libraries are well known in the art. Such homologue molecules may differ in their nucleotide sequences from those found in one or more of SEQ ID NO:1 through SEQ ID NO:32236 or complements thereof because complete complementarity is not needed for stable hybridization. The nucleic acid molecules of the present invention therefore also include molecules that, although capable of specifically hybridizing with the nucleic acid molecules may lack "complete complementarity."

Any of a variety of methods may be used to obtain one or more of the above-described nucleic acid molecules (Zamechik *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 83*:4143-4146 (1986), the entirety of which is herein incorporated by reference; Goodchild *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 85*:5507-5511 (1988), the entirety of which is herein incorporated by reference;

Wickstrom *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 85*:1028-1032 (1988), the entirety of which is herein incorporated by reference; Holt, *et al.*, *Molec. Cell. Biol. 8*:963-973 (1988), the entirety of which is herein incorporated by reference; Gerwirtz, *et al.*, *Science 242*:1303-1306 (1988), the entirety of which is herein incorporated by reference; Anfossi, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 86*:3379-3383 (1989), the entirety of which is herein incorporated by reference; Becker, *et al.*, *EMBO J. 8*:3685-3691 (1989); the entirety of which is herein incorporated by reference). Automated nucleic acid synthesizers may be employed for this purpose. In lieu of such synthesis, the disclosed nucleic acid molecules may be used to define a pair of primers that can be used with the polymerase chain reaction (Mullis, *et al.*, *Cold Spring Harbor Symp. Quant. Biol. 51*:263-273 (1986); Erlich *et al.*, EP 50,424; EP 84,796, EP 258,017, EP 237,362; Mullis, EP 201,184; Mullis *et al.*, US 4,683,202; Erlich, US 4,582,788; and Saiki, R. *et al.*, US 4,683,194, all of which are hereby incorporated by reference in their entirety) to amplify and obtain any desired nucleic acid molecule or fragment.

Promoter sequence(s) and other genetic elements including but not limited to transcriptional regulatory elements associated with one or more of the disclosed nucleic acid sequences can also be obtained using the disclosed nucleic acid sequences provided herein.

In one embodiment, such sequences are obtained by incubating EST nucleic acid molecules or preferably fragments thereof with members of genomic libraries (*e.g.* maize and soybean) and recovering clones that hybridize to the EST nucleic acid molecule or fragment thereof. In a second embodiment, methods of "chromosome walking," or inverse PCR may be used to obtain such sequences (Frohman, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 85*:8998-9002 (1988); Ohara, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 86*: 5673-5677 (1989); Pang *et al.*, *Biotechniques*, 22(6); 1046-1048 (1977); Huang *et al.*, *Methods Mol. Biol.* 69: 89-96 (1977); Hartl *et al.*, *Methods Mol. Biol.* 58: 293-301 (1996), all of which are hereby incorporated by reference in their entirety). In one embodiment, the disclosed nucleic acid molecules are used to identify cDNAs whose analogous genes contain promoters with desirable expression patterns. The nucleic acid molecules isolated from the library of the present invention are used to isolate

promoters of tissue-enhanced, tissue-specific, developmentally- or environmentally-regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences of these genes from genomic libraries, for example, using genomic screening methods and PCR techniques would result in the isolation of useful promoters and transcriptional regulatory elements. These methods are known to those of skill in the art and have been described (See for example Birren *et al.*, *Genome Analysis:Analyzing DNA*, 1, (1997), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., the entirety of which is herein incorporated by reference).

Promoters obtained utilizing the nucleic acid molecules of the present invention could also be modified to affect their control characteristics. Examples of such modifications would include but are not limited to enhancer sequences as reported by Kay *et al.*, *Science 236*:1299 (1987), herein incorporated by reference in its entirety. Such genetic elements could be used to enhance gene expression of new and existing traits for crop improvements.

The nucleic acid molecules of the present invention may be used to isolate promoters of tissue enhanced. tissue specific, cell-specific, cell -type, developmentally or environmentally regulated expression profiles. Isolation and functional analysis of the 5' flanking promoter sequences of these genes from genomic libraries, for example, using genomic screening methods and PCR techniques would result in the isolation of useful promoters and transcriptional regulatory elements. These methods are known to those of skill in the art and have been described (*See*, for example, *Birren et. al.*, *Genome Analysis*: *Analyzing DNA*, 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1997), the entirety of which is herein incorporated by reference). Promoters obtained utilizing the nucleic acid molecules of the present invention could also be modified to affect their control characteristics. Examples of such modifications would include but are not limited to enhancer sequences as reported by Kay, *et al* *Science 236*:1299 (1987), herein incorporated reference in its entirety. Such genetic elements could be used to enhance gene expression of new and existing traits for crop improvements.

In an aspect of the present invention, one or more of the nucleic molecules of the present invention are used to determine whether a plant (preferably maize) has a mutation affecting the level (*i.e.*, the concentration of mRNA in a sample, *etc.*) or pattern (*i.e.*, the kinetics of expression, rate of decomposition, stability profile, *etc.*) of the expression encoded in part or whole by one or more of the nucleic acid molecules of the present invention (collectively, the "Expression Response" of a cell or tissue). As used herein, the Expression Response manifested by a cell or tissue is said to be "altered" if it differs from the Expression Response of cells or tissues of plants not exhibiting the phenotype. To determine whether a Expression Response is altered, the Expression Response manifested by the cell or tissue of the plant exhibiting the phenotype is compared with that of a similar cell or tissue sample of a plant not exhibiting the phenotype. As will be appreciated, it is not necessary to re-determine the Expression Response of the cell or tissue sample of plants not exhibiting the phenotype each time such a comparison is made; rather, the Expression Response of a particular plant may be compared with previously obtained values of normal plants. As used herein, the phenotype of the organism is any of one or more characteristics of an organism (*e.g.* disease resistance, pest tolerance, environmental tolerance, male sterility, yield, quality improvements, *etc.*). A change in genotype or phenotype may be transient or permanent. Also as used herein, a tissue sample is any sample that comprises more than one cell. In a preferred aspect, a tissue sample comprises cells that share a common characteristic (*e.g.* derived from leaf, root, or pollen etc).

In one sub-aspect, such an analysis is conducted by determining the presence and/or identity of polymorphism(s) by one or more of the nucleic acid molecules of the present invention and more specifically, one or more of the EST nucleic acid molecules or fragments thereof which are associated with phenotype, or a predisposition to phenotype.

Any of a variety of molecules can be used to identify such polymorphism(s). In one embodiment, one or more of the EST nucleic acid molecules (or a sub-fragment thereof) may be employed as a marker nucleic acid molecule to identify such polymorphism(s). Alternatively, such polymorphisms can be detected through the use of a marker nucleic acid molecule or a

marker protein that is genetically linked to (*i.e.*, a polynucleotide that co-segregates with) such polymorphism(s).

In an alternative embodiment, such polymorphisms can be detected through the use of a marker nucleic acid molecule that is physically linked to such polymorphism(s). For this purpose, marker nucleic acid molecules comprising a nucleotide sequence of a polynucleotide located within 1 mb of the polymorphism(s), and more preferably within 100 kb of the polymorphism(s), and most preferably within 10 kb of the polymorphism(s) can be employed.

The genomes of animals and plants naturally undergo spontaneous mutation in the course of their continuing evolution (Gusella, *Ann. Rev. Biochem. 55*:831-854 (1986)). A "polymorphism" is a variation or difference in the sequence of the gene or its flanking regions that arises in some of the members of a species. The variant sequence and the "original" sequence co-exist in the species' population. In some instances, such co-existence is in stable or quasi-stable equilibrium.

A polymorphism is thus said to be "allelic," in that, due to the existence of the polymorphism, some members of a species may have the original sequence (*i.e.*, the original "allele") whereas other members may have the variant sequence (*i.e.*, the variant "allele"). In the simplest case, only one variant sequence may exist, and the polymorphism is thus said to be di-allelic. In other cases, the species' population may contain multiple alleles, and the polymorphism is termed tri-allelic, *etc.* A single gene may have multiple different unrelated polymorphisms. For example, it may have a di-allelic polymorphism at one site, and a multi-allelic polymorphism at another site.

The variation that defines the polymorphism may range from a single nucleotide variation to the insertion or deletion of extended regions within a gene. In some cases, the DNA sequence variations are in regions of the genome that are characterized by short tandem repeats (STRs) that include tandem di- or tri-nucleotide repeated motifs of nucleotides. Polymorphisms characterized by such tandem repeats are referred to as "variable number tandem repeat" ("VNTR") polymorphisms. VNTRs have been used in identity analysis (Weber, U.S. Patent

5,075,217; Armour, *et al.*, *FEBS Lett. 307*:113-115 (1992); Jones, *et al.*, *Eur. J. Haematol. 39*:144-147 (1987); Horn, *et al.*, PCT Application WO91/14003; Jeffreys, European Patent Application 370,719; Jeffreys, U.S. Patent 5,699,082; Jeffreys. *et al.*, *Amer. J. Hum. Genet. 39*:11-24 (1986); Jeffreys. *et al.*, *Nature 316*:76-79 (1985); Gray, *et al.*, *Proc. R. Acad. Soc. Lond. 243*:241-253 (1991); Moore, *et al.*, *Genomics 10*:654-660 (1991); Jeffreys, *et al.*, *Anim. Genet. 18*:1-15 (1987); Hillel, *et al.*, *Anim. Genet. 20*:145-155 (1989); Hillel, *et al.*, *Genet. 124*:783-789 (1990), all of which are herein incorporated by reference in their entirety).

The detection of polymorphic sites in a sample of DNA may be facilitated through the use of nucleic acid amplification methods. Such methods specifically increase the concentration of polynucleotides that span the polymorphic site, or include that site and sequences located either distal or proximal to it. Such amplified molecules can be readily detected by gel electrophoresis or other means.

The most preferred method of achieving such amplification employs the polymerase chain reaction ("PCR") (Mullis, *et al.*, *Cold Spring Harbor Symp. Quant. Biol. 51*:263-273 (1986); Erlich, *et al.*, European Patent Appln. 50,424; European Patent Appln. 84,796, European Patent Application 258,017, European Patent Appln. 237,362; Mullis, European Patent Appln. 201,184; Mullis, *et al.*, U.S. Patent No. 4,683,202; Erlich., U.S. Patent No. 4,582,788; and Saiki, *et al.*, U.S. Patent No. 4,683,194, all of which are herein incorporated by reference), using primer pairs that are capable of hybridizing to the proximal sequences that define a polymorphism in its double-stranded form.

In lieu of PCR, alternative methods, such as the "Ligase Chain Reaction" ("LCR") may be used (Barany, *Proc. Natl. Acad. Sci. (U.S.A.) 88*:189-193 (1991), the entirety of which is herein incorporated by reference). LCR uses two pairs of oligonucleotide probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependent ligase. As with PCR, the resulting products thus serve as a

template in subsequent cycles and an exponential amplification of the desired sequence is obtained.

LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a polymorphic site. In one embodiment, either oligonucleotide will be designed to include the actual polymorphic site of the polymorphism. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the polymorphic site present on the oligonucleotide. Alternatively, the oligonucleotides may be selected such that they do not include the polymorphic site (see, Segev, PCT Application WO 90/01069, the entirety of which is herein incorporated by reference).

The "Oligonucleotide Ligation Assay" ("OLA") may alternatively be employed (Landegren, *et al.*, *Science 241*:1077-1080 (1988), the entirety of which is herein incorporated by reference). The OLA protocol uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target. OLA, like LCR, is particularly suited for the detection of point mutations. Unlike LCR, however, OLA results in "linear" rather than exponential amplification of the target sequence.

Nickerson, *et al.* have described a nucleic acid detection assay that combines attributes of PCR and OLA (Nickerson, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 87*:8923-8927 (1990), the entirety of which is herein incorporated by reference). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA. In addition to requiring multiple, and separate, processing steps, one problem associated with such combinations is that they inherit all of the problems associated with PCR and OLA.

Schemes based on ligation of two (or more) oligonucleotides in the presence of nucleic acid having the sequence of the resulting "di-oligonucleotide", thereby amplifying the di-oligonucleotide, are also known (Wu, *et al.*, *Genomics 4*:560 (1989), the entirety of which is herein incorporated by reference), and may be readily adapted to the purposes of the present invention.

Other known nucleic acid amplification procedures, such as allele-specific oligomers, branched DNA technology, transcription-based amplification systems, or isothermal amplification methods may also be used to amplify and analyze such polymorphisms (Malek, *et al.*, U.S. Patent 5,130,238; Davey, *et al.*, European Patent Application 329,822; Schuster *et al.*, U.S. Patent 5,169,766; Miller, *et al.*, PCT Application WO 89/06700; Kwoh, *et al.*, *Proc. Natl. Acad. Sci.* (U.S.A.) *86*:1173-1177 (1989); Gingeras, *et al.*, PCT Application WO 88/10315; Walker, *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 89*:392-396 (1992), all of which are herein incorporated by reference in their entirety).

The identification of a polymorphism can be determined in a variety of ways. By correlating the presence or absence of it in a plant with the presence or absence of a phenotype, it is possible to predict the phenotype of that plant. If a polymorphism creates or destroys a restriction endonuclease cleavage site, or if it results in the loss or insertion of DNA (*e.g.*, a VNTR polymorphism), it will alter the size or profile of the DNA fragments that are generated by digestion with that restriction endonuclease. As such, individuals that possess a variant sequence can be distinguished from those having the original sequence by restriction fragment analysis. Polymorphisms that can be identified in this manner are termed "restriction fragment length polymorphisms" ("RFLPs"). RFLPs have been widely used in human and plant genetic analyses (Glassberg, UK Patent Application 2135774; Skolnick, *et al.*, *Cytogen. Cell Genet. 32*:58-67 (1982); Botstein, *et al.*, *Ann. J. Hum. Genet. 32*:314-331 (1980); Fischer, *et al.* (PCT Application WO90/13668); Uhlen, PCT Application WO90/11369).

Polymorphisms can also be identified by Single Strand Conformation Polymorphism (SSCP) analysis. The SSCP technique is a method capable of identifying most sequence variations in a single strand of DNA, typically between 150 and 250 nucleotides in length (Elles, *Methods in Molecular Medicine: Molecular Diagnosis of Genetic Diseases*, Humana Press (1996), the entirety of which is herein incorporated by reference); Orita *et al.*, *Genomics 5:* 874-879 (1989), the entirety of which is herein incorporated by reference). Under denaturing conditions a single strand of DNA will adopt a conformation that is uniquely dependent on its

sequence conformation. This conformation usually will be different, even if only a single base is changed. Most conformations have been reported to alter the physical configuration or size sufficiently to be detectable by electrophoresis. A number of protocols have been described for SSCP including, but not limited to Lee *et al., Anal. Biochem. 205:* 289-293 (1992), the entirety of which is herein incorporated by reference; Suzuki *et al., Anal. Biochem. 192:* 82-84 (1991), the entirety of which is herein incorporated by reference; Lo *et al., Nucleic Acids Research 20:* 1005-1009 (1992), the entirety of which is herein incorporated by reference; Sarkar *et al., Genomics 13:* 441-443 (1992), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by SSCP analysis.

Polymorphisms may also be found using a DNA fingerprinting technique called amplified fragment length polymorphism (AFLP), which is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA to profile that DNA. Vos, *et al., Nucleic Acids Res. 23*:4407-4414 (1995), the entirety of which is herein incorporated by reference. This method allows for the specific co-amplification of high numbers of restriction fragments, which can be visualized by PCR without knowledge of the nucleic acid sequence.

AFLP employs basically three steps. Initially, a sample of genomic DNA is cut with restriction enzymes and oligonucleotide adapters are ligated to the restriction fragments of the DNA. The restriction fragments are then amplified using PCR by using the adapter and restriction sequence as target sites for primer annealing. The selective amplification is achieved by the use of primers that extend into the restriction fragments, amplifying only those fragments in which the primer extensions match the nucleotide flanking the restriction sites. These amplified fragments are then visualized on a denaturing polyacrylamide gel.

AFLP analysis has been performed on *Salix* (Beismann, *et al., Mol. Ecol. 6*:989-993 (1997), the entirety of which is herein incorporated by reference); *Acinetobacter* (Janssen, *et al., Int. J. Syst. Bacteriol 47*:1179-1187 (1997), the entirety of which is herein incorporated by reference), *Aeromonas popoffi* (Huys, *et al., Int. J. Syst. Bacteriol. 47*:1165-1171 (1997), the

entirety of which is herein incorporated by reference), rice (McCouch, *et al.*, *Plant Mol. Biol.* *35*:89-99 (1997), the entirety of which is herein incorporated by reference); Nandi, *et al.*, *Mol. Gen. Genet.* *255*:1-8 (1997); Cho, *et al.*, *Genome* *39*:373-378 (1996), herein incorporated by reference), barley (*Hordeum vulgare*)(Simons, *et al.*, *Genomics* *44*:61-70 (1997), the entirety of which is herein incorporated by reference; Waugh, *et al.*, *Mol. Gen. Genet.* *255*:311-321 (1997), the entirety of which is herein incorporated by reference; Qi, *et al.*, *Mol. Gen Genet.* *254*:330-336 (1997), the entirety of which is herein incorporated by reference; Becker, *et al.*, *Mol. Gen. Genet.* *249*:65-73 (1995), the entirety of which is herein incorporated by reference), potato (Van der Voort, *et al.*, *Mol. Gen. Genet.* *255*:438-447 (1997), the entirety of which is herein incorporated by reference; Meksem, *et al.*, *Mol. Gen. Genet.* *249*:74-81 (1995), the entirety of which is herein incorporated by reference), *Phytophthora infestans* (Van der Lee, *et al.*, *Fungal Genet. Biol.* *21*:278-291 (1997), the entirety of which is herein incorporated by reference), *Bacillus anthracis* (Keim, *et al.*, *J. Bacteriol.* *179*:818-824 (1997)), *Astragalus cremnophylax* (Travis, *et al.*, *Mol. Ecol.* *5*:735-745 (1996), the entirety of which is herein incorporated by reference), *Arabidopsis* (Cnops, *et al.*, *Mol. Gen. Genet.* *253*:32-41 (1996), the entirety of which is herein incorporated by reference), *Escherichia coli* (Lin, *et al.*, *Nucleic Acids Res.* *24*:3649-3650 (1996), the entirety of which is herein incorporated by reference), *Aeromonas* (Huys, *et al.*, *Int. J. Syst. Bacteriol.* *46*:572-580 (1996), the entirety of which is herein incorporated by reference), nematode (Folkertsma, *et al.*, *Mol. Plant Microbe Interact.* *9*:47-54 (1996), the entirety of which is herein incorporated by reference), tomato (Thomas, *et al.*, *Plant J.* *8*:785-794 (1995), the entirety of which is herein incorporated by reference), and human (Latorra, *et al.*, *PCR Methods Appl.* *3*:351-358 (1994)). AFLP analysis has also been used for fingerprinting mRNA (Money, *et al.*, *Nucleic Acids Res.* *24*:2616-2617 (1996), the entirety of which is herein incorporated by reference; Bachem, *et al.*, *Plant J.* *9*:745-753 (1996), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by AFLP analysis for fingerprinting mRNA.

Polymorphisms may also be found using random amplified polymorphic DNA (RAPD) (Williams *et al., Nucl. Acids Res. 18*: 6531-6535 (1990), the entirety of which is herein incorporated by reference) and cleaveable amplified polymorphic sequences (CAPS) (Lyamichev *et al., Science 260*: 778-783 (1993), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acids of the present invention, may be utilized as markers or probes to detect polymorphisms by RAPD or CAPS analysis.

Polymorphisms are useful, through linkage analysis, to define the genetic distances or physical distances between polymorphic traits. A physical map or ordered array of genomic DNA fragments in the desired region containing the gene may be used to characterize and isolate genes corresponding to desirable traits. For this purpose, yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), and cosmids are appropriate vectors for cloning large segments of DNA molecules. Although fewer clones are needed to make a contig for a specific genomic region by using YACs (Agyare *et al., Genome Res. 7*: 1-9 (1997), the entirety of which is herein incorporated by reference; James *et al., Genomics 32*: 425-430 (1996), the entirety of which is herein incorporated by reference), chimerism in the inserted DNA fragment can arise. Cosmids are convenient for handling smaller-size DNA molecules and may be used for transformation in developing transgenic plants. BACs also carry DNA fragments and are less prone to chimerism.

Through genetic mapping, a fine scale linkage map can be developed using DNA markers and, then, a genomic DNA library of large-sized fragments can be screened with molecular markers linked to the desired trait. Molecular markers are advantageous for agronomic traits that are otherwise difficult to tag, such as resistance to pathogens, insects and nematodes, tolerance to abiotic stress, quality parameters and quantitative traits such as high yield potential.

The essential requirements for marker-assisted selection in a plant breeding program are: (1) the marker(s) should co-segregate or be closely linked with the desired trait; (2) an efficient means of screening large populations for the molecular marker(s) should be available; and (3) the

screening technique should have high reproducibility across laboratories and preferably be economical to use and be user-friendly.

The genetic linkage of marker molecules can be established by a gene mapping model such as, without limitation, the flanking marker model reported by Lander and Botstein, *Genetics 121*:185-199 (1989) and the interval mapping, based on maximum likelihood methods described by Lander and Botstein, *Genetics 121*:185-199 (1989) and implemented in the software package MAPMAKER/QTL (Lincoln and Lander, *Mapping Genes Controlling Quantitative Traits Using MAPMAKER/QTL*, Whitehead Institute for Biomedical Research, Massachusetts, (1990). Additional software includes Qgene, Version 2.23 (1996), Department of Plant Breeding and Biometry, 266 Emerson Hall, Cornell University, Ithaca, NY, the manual of which is herein incorporated by reference in its entirety). Use of Qgene software is a particularly preferred approach.

A maximum likelihood estimate (MLE) for the presence of a marker is calculated, together with an MLE assuming no QTL effect, to avoid false positives. A $\log_{10}$ of an odds ratio (LOD) is then calculated as: LOD = $\log_{10}$ (MLE for the presence of a QTL/MLE given no linked QTL).

The LOD score essentially indicates how much more likely the data are to have arisen assuming the presence of a QTL than in its absence. The LOD threshold value for avoiding a false positive with a given confidence, say 95%, depends on the number of markers and the length of the genome. Graphs indicating LOD thresholds are set forth in Lander and Botstein, *Genetics 121*:185-199 (1989) the entirety of which is herein incorporated by reference and further described by Arús and Moreno-González , *Plant Breeding*, Hayward *et al.*, (eds.) Chapman & Hall, London, pp. 314-331 (1993), the entirety of which is herein incorporated by reference.

Additional models can be used. Many modifications and alternative approaches to interval mapping have been reported, including the use of non-parametric methods (Kruglyak and Lander, *Genetics 139*:1421-1428 (1995), the entirety of which is herein incorporated by

reference). Multiple regression methods or models can be also be used, in which the trait is regressed on a large number of markers (Jansen, *Biometrics in Plant Breeding*, van Oijen and Jansen (eds.), Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp. 116-124 (1994); Weber and Wricke, *Advances in Plant Breeding*, Blackwell, Berlin, 16 (1994), both of which is herein incorporated by reference in their entirety). Procedures combining interval mapping with regression analysis, whereby the phenotype is regressed onto a single putative QTL at a given marker interval and at the same time onto a number of markers that serve as 'cofactors,' have been reported by Jansen and Stam, *Genetics 136*:1447-1455 (1994), the entirety of which is herein incorporated by reference and Zeng, *Genetics 136*:1457-1468 (1994) the entirety of which is herein incorporated by reference. Generally, the use of cofactors reduces the bias and sampling error of the estimated QTL positions (Utz and Melchinger, *Biometrics in Plant Breeding*, van Oijen and Jansen (eds.) Proceedings of the Ninth Meeting of the Eucarpia Section Biometrics in Plant Breeding, The Netherlands, pp.195-204 (1994), the entirety of which is herein incorporated by reference, thereby improving the precision and efficiency of QTL mapping (Zeng, *Genetics 136*:1457-1468 (1994)). These models can be extended to multi-environment experiments to analyze genotype-environment interactions (Jansen *et al., Theo. Appl. Genet. 91*:33-37 (1995), the entirety of which is herein incorporated by reference).

Selection of an appropriate mapping population is important to map construction. The choice of an appropriate mapping population depends on the type of marker systems employed (Tanksley *et al., Molecular mapping plant chromosomes. Chromosome structure and function: Impact of new concepts*, Gustafson and Appels (eds.), Plenum Press, New York, pp. 157-173 (1988), the entirety of which is herein incorporated by reference). Consideration must be given to the source of parents (adapted vs. exotic) used in the mapping population. Chromosome pairing and recombination rates can be severely disturbed (suppressed) in wide crosses (adapted x exotic) and generally yield greatly reduced linkage distances. Wide crosses will usually

provide segregating populations with a relatively large array of polymorphisms when compared to progeny in a narrow cross (adapted x adapted).

An F_2 population is the first generation of selfing after the hybrid seed is produced. Usually a single F_1 plant is selfed to generate a population segregating for all the genes in Mendelian (1:2:1) fashion. Maximum genetic information is obtained from a completely classified F_2 population using a codominant marker system (Mather, *Measurement of Linkage in Heredity*, Methuen and Co., (1938), the entirety of which is herein incorporated by reference). In the case of dominant markers, progeny tests (*e.g.* F_3, BCF_2) are required to identify the heterozygotes, thus making it equivalent to a completely classified F_2 population. However, this procedure is often prohibitive because of the cost and time involved in progeny testing. Progeny testing of F_2 individuals is often used in map construction where phenotypes do not consistently reflect genotype (*e.g.* disease resistance) or where trait expression is controlled by a QTL. Segregation data from progeny test populations (*e.g.* F_3 or BCF_2) can be used in map construction. Marker-assisted selection can then be applied to cross progeny based on marker-trait map associations (F_2, F_3), where linkage groups have not been completely disassociated by recombination events (*i.e.*, maximum disequilibrium).

Recombinant inbred lines (RIL) (genetically related lines; usually >F_5, developed from continuously selfing F_2 lines towards homozygosity) can be used as a mapping population. Information obtained from dominant markers can be maximized by using RIL because all loci are homozygous or nearly so. Under conditions of tight linkage (*i.e.*, about <10% recombination), dominant and co-dominant markers evaluated in RIL populations provide more information per individual than either marker type in backcross populations (Reiter *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 89*:1477-1481 (1992), the entirety of which is herein incorporated by reference). However, as the distance between markers becomes larger (*i.e.*, loci become more independent), the information in RIL populations decreases dramatically when compared to codominant markers.

Backcross populations (*e.g.*, generated from a cross between a successful variety (recurrent parent) and another variety (donor parent) carrying a trait not present in the former) can be utilized as a mapping population. A series of backcrosses to the recurrent parent can be made to recover most of its desirable traits. Thus a population is created consisting of individuals nearly like the recurrent parent but each individual carries varying amounts or mosaic of genomic regions from the donor parent. Backcross populations can be useful for mapping dominant markers if all loci in the recurrent parent are homozygous and the donor and recurrent parent have contrasting polymorphic marker alleles (Reiter *et al., Proc. Natl. Acad. Sci. (U.S.A.)* *89*:1477-1481 (1992)). Information obtained from backcross populations using either codominant or dominant markers is less than that obtained from F_2 populations because one, rather than two, recombinant gametes are sampled per plant. Backcross populations, however, are more informative (at low marker saturation) when compared to RILs as the distance between linked loci increases in RIL populations (*i.e.* about 15% recombination). Increased recombination can be beneficial for resolution of tight linkages, but may be undesirable in the construction of maps with low marker saturation.

Near-isogenic lines (NIL) created by many backcrosses to produce an array of individuals that are nearly identical in genetic composition except for the trait or genomic region under interrogation can be used as a mapping population. In mapping with NILs, only a portion of the polymorphic loci are expected to map to a selected region.

Bulk segregant analysis (BSA) is a method developed for the rapid identification of linkage between markers and traits of interest (Michelmore *et al., Proc. Natl. Acad. Sci. (U.S.A.)* *88*:9828-9832 (1991), the entirety of which is herein incorporated by reference). In BSA, two bulked DNA samples are drawn from a segregating population originating from a single cross. These bulks contain individuals that are identical for a particular trait (resistant or susceptible to particular disease) or genomic region but arbitrary at unlinked regions (*i.e.* heterozygous). Regions unlinked to the target region will not differ between the bulked samples of many individuals in BSA.

It is understood that one or more of the nucleic acid molecules of the present invention may be used as molecular markers. It is also understood that one or more of the protein molecules of the present invention may be used as molecular markers.

In accordance with this aspect of the present invention, a sample nucleic acid is obtained from plants cells or tissues. Any source of nucleic acid may be used. Preferably, the nucleic acid is genomic DNA. The nucleic acid is subjected to restriction endonuclease digestion. For example, one or more EST nucleic acid molecule or fragment thereof can be used as a probe in accordance with the above-described polymorphic methods. The polymorphism obtained in this approach can then be cloned to identify the mutation at the coding region which alters the protein's structure or regulatory region of the gene which affects its expression level.

In one aspect of the present invention, an evaluation can be conducted to determine whether a particular mRNA molecule is present. One or more of the nucleic acid molecules of the present invention, preferably one or more of the EST nucleic acid molecules of the present invention are utilized to detect the presence or quantity of the mRNA species. Such molecules are then incubated with cell or tissue extracts of a plant under conditions sufficient to permit nucleic acid hybridization. The detection of double-stranded probe-mRNA hybrid molecules is indicative of the presence of the mRNA; the amount of such hybrid formed is proportional to the amount of mRNA. Thus, such probes may be used to ascertain the level and extent of the mRNA production in a plant's cells or tissues. Such nucleic acid hybridization may be conducted under quantitative conditions (thereby providing a numerical value of the amount of the mRNA present). Alternatively, the assay may be conducted as a qualitative assay that indicates either that the mRNA is present, or that its level exceeds a user set, predefined value.

A principle of *in situ* hybridization is that a labeled, single-stranded nucleic acid probe will hybridize to a complementary strand of cellular DNA or RNA and, under the appropriate conditions, these molecules will form a stable hybrid. When nucleic acid hybridization is combined with histological techniques, specific DNA or RNA sequences can be identified within a single cell. An advantage of *in situ* hybridization over more conventional techniques for the

detection of nucleic acids is that it allows an investigator to determine the precise spatial population (Angerer *et al., Dev. Biol. 101*: 477-484 (1984), the entirety of which is herein incorporated by reference; Angerer *et al., Dev. Biol. 112*: 157-166 (1985), the entirety of which is herein incorporated by reference; Dixon *et al., EMBO J. 10*: 1317-1324 (1991), the entirety of which is herein incorporated by reference). *In situ* hybridization may be used to measure the steady-state level of RNA accumulation. It is a sensitive technique and RNA sequences present in as few as 5-10 copies per cell can be detected (Hardin *et al., J. Mol. Biol. 202*: 417-431.(1989), the entirety of which is herein incorporated by reference). A number of protocols have been devised for *in situ* hybridization, each with tissue preparation, hybridization, and washing conditions (Meyerowitz, *Plant Mol. Biol. Rep. 5*: 242-250 (1987), the entirety of which is herein incorporated by reference; Cox and Goldberg, In: *Plant Molecular Biology: A Practical Approach* (ed. C.H. Shaw), pp. 1-35. IRL Press, Oxford (1988), the entirety of which is herein incorporated by reference; Raikhel *et al., In situ RNA hybridization in plant tissues.* In *Plant Molecular Biology Manual*, vol. B9: 1-32. Kluwer Academic Publisher, Dordrecht, Belgium (1989), the entirety of which is herein incorporated by reference).

In situ hybridization also allows for the localization of proteins within a tissue or cell (Wilkinson, *In Situ Hybridization*, Oxford University Press, Oxford (1992), the entirety of which is herein incorporated by reference; Langdale, *In Situ Hybridization* 165-179 In: *The Maize Handbook*, eds. Freeling and Walbot, Springer-Verlag, New York (1994), the entirety of which is herein incorporated by reference). It is understood that one or more of the molecules of the present invention, preferably one or more of the EST nucleic acid molecules of the present invention or one or more of the antibodies of the present invention may be utilized to detect the level or pattern of a protein or fragment thereof by *in situ* hybridization.

Fluorescent *in situ* hybridization also enables the localization of a particular DNA sequence along a chromosome which is useful, among other uses, for gene mapping, following chromosomes in hybrid lines or detecting chromosomes with translocations, transversions or deletions. *In situ* hybridization has been used to identify chromosomes in several plant species

(Griffor *et al., Plant Mol. Biol. 17*: 101-109 (1991), the entirety of which is herein incorporated by reference; Gustafson *et al., Proc. Nat'l. Acad. Sci. (U.S.A). 87*: 1899-1902 (1990), herein incorporated by reference; Mukai and Gill, *Genome 34*: 448-452. (1991); Schwarzacher and Heslop-Harrison, *Genome 34*: 317-323 (1991); Wang *et al., Jpn. J. Genet. 66*: 313-316 (1991), the entirety of which is herein incorporated by reference; Parra and Windle, *Nature Genetics, 5*: 17-21 (1993), the entirety of which is herein incorporated by reference). It is understood that the nucleic acid molecules of the present invention may be used as probes or markers to localize sequences along a chromosome.

It is also understood that one or more of the molecules of the present invention, preferably one or more of the EST nucleic acid molecules of the present invention or one or more of the antibodies of the present invention may be utilized to detect the expression level or pattern of a protein or mRNA thereof by *in situ* hybridization.

Another method to localize the expression of a molecule is tissue printing. Tissue printing provides a way to screen, at the same time on the same membrane many tissue sections from different plants or different developmental stages. Tissue-printing procedures utilize films designed to immobilize proteins and nucleic acids. In essence, a freshly cut section of an organ is pressed gently onto nitrocellulose paper, nylon membrane or polyvinylidene difluoride membrane. Such membranes are commercially available (*e.g.* Millipore, Bedford, Massachusetts). The contents of the cut cell transfer onto the membrane, and the molecules are immobilized to the membrane. The immobilized molecules form a latent print that can be visualized with appropriate probes. When a plant tissue print is made on nitrocellulose paper, the cell walls leave a physical print that makes the anatomy visible without further treatment (Varner and Taylor, *Plant Physiol. 91*: 31-33 (1989), the entirety of which is herein incorporated by reference).

Tissue printing on substrate films is described by Daoust, *Exp. Cell Res. 12*: 203-211 (1957), the entirety of which is herein incorporated by reference, who detected amylase, protease, ribonuclease, and deoxyribonuclease in animal tissues using starch, gelatin, and agar films.

These techniques can be applied to plant tissues (Yomo and Taylor, *Planta 112*:35-43 (1973); Harris and Chrispeels, *Plant Physiol.* 56: 292-299 (1975). Advances in membrane technology have increased the range of applications of Daoust's tissue-printing techniques allowing (Cassab and Varner, *J. Cell. Biol.* 105: 2581-2588 (1987), the entirety of which is herein incorporated by reference; the histochemical localization of various plant enzymes and deoxyribonuclease on nitrocellulose paper and nylon (Spruce *et al.*, *Phytochemistry, 26:* 2901-2903 (1987), the entirety of which is herein incorporated by reference; Barres *et al. Neuron 5:* 527-544 (1990), the entirety of which is herein incorporated by reference; the entirety of which is herein incorporated by reference; Reid and Pont-Lezica, *Tissue Printing: Tools for the Study of Anatomy, Histochemistry, and Gene Expression,* Academic Press, New York, New York (1992), the entirety of which is herein incorporated by reference; Reid *et al. Plant Physiol. 93:* 160-165 (1990), herein incorporate by reference; Ye *et al. Plant J. 1:* 175-183 (1991), the entirety of which is herein incorporated by reference).

It is understood that one or more of the molecules of the present invention, preferably one or more of the EST nucleic acid molecules of the present invention or one or more of the antibodies of the present invention may be utilized to detect the presence or quantity of a protein by tissue printing.

Further, it is also understood that any of the nucleic acid molecules of the present invention may be used as marker nucleic acids and or probes in connection with methods that require probes or marker nucleic acids. As used herein, a probe is an agent that is utilized to determine an attribute or feature (*e.g.* presence or absence, location, correlation, *etc.*) or a molecule, cell, tissue or plant. As used herein, a marker nucleic acid is a nucleic acid molecule that is utilized to determine an attribute or feature (*e.g.*, presence or absence, location, correlation, *etc.*) or a molecule, cell, tissue or plant.

A microarray-based method for high-throughput monitoring of plant gene expression may be utilized to measure gene-specific hybridization targets. This 'chip'-based approach involves using microarrays of nucleic acid molecules as gene-specific hybridization targets to

quantitatively measure expression of the corresponding plant genes (Schena *et al.*, *Science 270*: 467-470 (1995), the entirety of which is herein incorporated by reference; Shalon, Ph.D. Thesis. Stanford University (1996), the entirety of which is herein incorporated by reference). Every nucleotide in a large sequence can be queried at the same time. Hybridization can be used to efficiently analyze large amounts of nucleotide sequence.

Several microarray methods have been described. One method compares the sequences to be analyzed by hybridization to a set of oligonucleotides representing all possible subsequences (Bains and Smith, *J. Theor. Biol.* 135: 303 (1989), the entirety of which is herein incorporated by reference). A second method hybridizes the sample to an array of oligonucleotide probes. An array consisting of oligonucleotides complementary to subsequences of a target sequence can be used to determine the identity of a target sequence, measure its amount, and detect differences between the target and a reference sequence. Nucleic acid molecules microarrays may also be screened with protein molecules or fragments thereof to determine nucleic acid molecules that specifically bind protein molecules or fragments thereof.

The microarray approach may be used with polypeptide targets (U.S. Patent No. 5,445,934; U.S. Patent No: 5,143,854; U.S. Patent No. 5,079,600; U.S. Patent No. 4,923,901, all of which are herein incorporated by reference in their entirety). Essentially, polypeptides are synthesized on a substrate (microarray) and these polypeptides can be screened with either protein molecules or fragments thereof or nucleic acid molecules in order to screen for either protein molecules or fragments thereof or nucleic acid molecules that specifically bind the target polypeptides. Implementation of these techniques rely on recently developed combinatorial technologies to generate any ordered array of a large number of oligonucleotide probes (Fodor *et al.*, *Science 251*:767-773 (1991), the entirety of which is herein incorporated by reference).

It is understood that one or more of the molecules of the present invention, preferably one or more of the nucleic acid molecules or protein molecules or fragments thereof of the present invention may be utilized in a microarray based method.

In a preferred embodiment of the present invention microarrays may be prepared that comprise nucleic acid molecules where preferably at least 10%, preferably at least 25%, more preferably at least 50% and even more preferably at least 75%, 80%, 85%, 90% or 95% of the nucleic acid molecules located on that array are selected from the group of nucleic acid molecules that specifically hybridize to one or more nucleic acid molecule having a nucleic acid sequence selected from the group of SEQ ID NO: 1 through SEQ ID NO: 32236 or complement thereof or fragments of either.

A particular preferred microarray embodiment of the present invention is a microarray comprising nucleic acid molecules encoding genes or fragments thereof that are homologues of known genes or nucleic acid molecules that comprise genes or fragment thereof that elicit only limited or no matches to known genes. A further preferred microarray embodiment of the present invention is a microarray comprising nucleic acid molecules having genes or fragments thereof that are homologues of known genes and nucleic acid molecules that comprise genes or fragment thereof that elicit only limited or no matches to known genes. Site-directed mutagenesis may be utilized to modify nucleic acid sequences, particularly as it is a technique that allows one or more of the amino acids encoded by a nucleic acid molecule to be altered (*e.g.* a threonine to be replaced by a methionine). Three basic methods for site-directed mutagenesis are often employed. These are cassette mutagenesis (Wells *et al.*, *Gene 34*:315-23 (1985), the entirety of which is herein incorporated by reference), primer extension (Gilliam *et al.*, *Gene 12*:129-137 (1980), the entirety of which is herein incorporated by reference); Zoller and Smith, *Methods Enzymol. 100*:468-500 (1983), the entirety of which is herein incorporated by reference; and Dalbadie-McFarland *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 79*:6409-6413 (1982), the entirety of which is herein incorporated by reference) and methods based upon PCR (Scharf *et al.*, *Science 233*:1076-1078 (1986), the entirety of which is herein incorporated by reference; Higuchi *et al.*, *Nucleic Acids Res. 16*:7351-7367 (1988), the entirety of which is herein incorporated by reference). Site-directed mutagenesis approaches are also described in European Patent 0 385 962, the entirety of which is herein incorporated by reference, European Patent 0 359 472, the

entirety of which is herein incorporated by reference, and PCT Patent Application WO 93/07278, the entirety of which is herein incorporated by reference.

Site-directed mutagenesis strategies have been applied to plants for both *in vitro* as well as *in vivo* site-directed mutagenesis (Lanz *et al., J. Biol. Chem.* 266:9971-6 (1991); Kovgan and Zhdanov, *Biotekhnologiya 5*:148-154; No. 207160n, Chemical Abstracts 110:225 (1989); Ge *et al., Proc. Natl. Acad. Sci. (U.S.A.) 86*:4037-4041 (1989); Zhu *et al., J. Biol. Chem. 271*:18494-18498 (1996); Chu *et al., Biochemistry 33*:6150-6157 (1994); Small *et al., EMBO J. 11*:1291-1296 (1992); Cho *et al., Mol. Biotechnol. 8*:13-16 (1997), Kita *et al., J. Biol. Chem. 271*:26529-26535 (1996); Jin *et al., Mol. Microbiol. 7*:555-562 (1993); Hatfield and Vierstra, *J. Biol. Chem. 267*:14799-14803 (1992); Zhao *et al., Biochemistry 31*:5093-5099 (1992), all of which are herein incorporated by reference in their entirety).

Any of the nucleic acid molecules of the present invention may either be modified by site-directed mutagenesis or used as, for example, nucleic acid molecules that are used to target other nucleic acid molecules for modification. It is understood that mutants with more than one altered nucleotide can be constructed using techniques that practitioners skilled in the art are familiar with such as isolating restriction fragments and ligating such fragments into an expression vector (*see*, for example, Sambrook *et al., Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989)).

Sequence-specific DNA-binding proteins play a role in the regulation of transcription. The isolation of recombinant cDNAs encoding these proteins facilitates the biochemical analysis of their structural and functional properties. Genes encoding such DNA-binding proteins have been isolated using classical genetics (Vollbrecht *et al., Nature 350*: 241-243 (1991), the entirety of which is herein incorporated by reference) and molecular biochemical approaches, including the screening of recombinant cDNA libraries with antibodies (Landschulz *et al., Genes Dev. 2*: 786-800 (1988), the entirety of which is herein incorporated by reference) or DNA probes (Bodner *et al., Cell 55*: 505-518 (1988), the entirety of which is herein incorporated by reference). In addition, an *in situ* screening procedure has been used and has facilitated the

isolation of sequence-specific DNA-binding proteins from various plant species (Gilmartin *et al.*, *Plant Cell 4*: 839-849 (1992), the entirety of which is herein incorporated by reference; Schindler *et al.*, *EMBO J. 11*: 1261-1273 (1992) the entirety of which is herein incorporated by reference). An *in situ* screening protocol does not require the purification of the protein of interest (Vinson *et al.*, *Genes Dev. 2*: 801-806 (1988), the entirety of which is herein incorporated by reference; Singh *et al.*, *Cell 52*: 415-423 (1988), the entirety of which is herein incorporated by reference).

Steps may be employed to characterize DNA-protein interactions. The first is to identify promoter fragments that interact with DNA-binding proteins, to titrate binding activity, to determine the specificity of binding, and to determine whether a given DNA-binding activity can interact with related DNA sequences (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). Electrophoretic mobility-shift assay is a widely used assay. The assay provides a simple, rapid, and sensitive method for detecting DNA-binding proteins based on the observation that the mobility of a DNA fragment through a nondenaturing, low-ionic strength polyacrylamide gel is retarded upon association with a DNA-binding protein (Fried and Crother, *Nucleic Acids Res. 9*: 6505-6525 (1981), the entirety of which is herein incorporated by reference). When one or more specific binding activities have been identified, the exact sequence of the DNA bound by the protein may be determined. Several procedures for characterizing protein/DNA-binding sites are used, including methylation and ethylation interference assays (Maxam and Gilbert, *Methods Enzymol. 65*: 499-560 (1980), the entirety of which is herein incorporated by reference; Wissman and Hillen, *Methods Enzymol. 208*: 365-379 (1991), the entirety of which is herein incorporated by reference) and footprinting techniques employing DNase I (Galas and Schmitz, *Nucleic Acids Res. 5*: 3157-3170 (1978), the entirety of which is herein incorporated by reference), 1,10-phenanthroline-copper ion methods (Sigman *et al.*, *Methods Enzymol. 208*: 365-379 (1991), the entirety of which is herein incorporated by reference) or hydroxyl radical methods (Dixon *et al.*, *Methods Enzymol. 208*: 380-413 (1991), the entirety of which is herein incorporated by reference). It is understood that one or more of the nucleic acid molecules of the

present invention, preferably one or more of the EST nucleic acid molecules of the present invention may be utilized to identify a protein or fragment thereof that specifically binds to a nucleic acid molecule of the present invention. It is also understood that one or more of the protein molecules or fragments thereof of the present invention may be utilized to identify a nucleic acid molecule that specifically binds to it.

The two-hybrid system is based on the fact that many cellular functions are carried out by proteins that interact (physically) with one another. Two-hybrid systems have been used to probe the function of new proteins (Chien *et al., Proc. Natl. Acad. Sci. (U.S.A.) 88*: 9578-9582 (1991) the entirety of which is herein incorporated by reference; Durfee *et al., Genes Dev. 7*: 555-569 (1993) the entirety of which is herein incorporated by reference; Choi *et al., Cell 78*: 499-512 (1994), the entirety of which is herein incorporated by reference; Kranz *et al., Genes Dev. 8:* 313-327 (1994), the entirety of which is herein incorporated by reference).

Interaction mating techniques have facilitated a number of two-hybrid studies of protein-protein interaction. Interaction mating has been used to examine interactions between small sets of tens of proteins (Finley and Brent, *Proc. Natl. Acad. Sci. (U.S.A.) 91*: 12098-12984 (1994), the entirety of which is herein incorporated by reference), larger sets of hundreds of proteins, (Bendixen *et al., Nucl. Acids Res. 22*: 1778-1779 (1994), the entirety of which is herein incorporated by reference) and to comprehensively map proteins encoded by a small genome (Bartel *et al., Nature Genetics 12*: 72-77 (1996), the entirety of which is herein incorporated by reference). This technique utilizes proteins fused to the DNA-binding domain and proteins fused to the activation domain. They are expressed in two different haploid yeast strains of opposite mating type, and the strains are mated to determine if the two proteins interact. Mating occurs when haploid yeast strains come into contact and result in the fusion of the two haploids into a diploid yeast strain. An interaction can be determined by the activation of a two-hybrid reporter gene in the diploid strain. The primary advantage of this technique is that it reduces the number of yeast transformations needed to test individual interactions. It is understood that the protein-protein interactions of protein or fragments thereof of the present invention may be investigated

using the two-hybrid system and that any of the nucleic acid molecules of the present invention that encode such proteins or fragments thereof may be used to transform yeast in the two-hybrid system.

Synechocystis 6803 is a photosynthetic Cyanobacterium capable of oxygenic photosynthesis as well as heterotrophic growth in the absence of light. The entire genome has been sequenced, and it is reported to have a circular genome size of 3.57 Mbp containing 3168 potential open reading frames. Open reading frames (ORFs) were identified based upon their homology to other reported ORFs and by using ORF identification computer programs. Sixteen hundred potential ORFs were assigned based on their homology to previously identified ORFs. Of these 1600 ORFs, 145 were identical to reported ORFs (Kaneko *et al., DNA Research 3*:109-36 (1996), herein incorporated by reference in its entirety).

Several prokaryote promoters have been used in *Synechocystis* to express heterologous genes including the tac, lac, and lambda phage promoters (Bryant (ed.), *The Molecular Biology of Cyanobacteria*, Kluwer Academic Publishers, (1994); Ferino and Chauvat, *Gene 84*:257-266 (1989), both of which are herein incorporated by reference in their entirety). Several bacterial origins of replication such as RSF1010 and ACYC are reported to replicate in *Synechocystis* (Mermet-Bouvier and Chauvat, *Current Microbiology 28*:145-148 (1994); Kuhlemeier *et al., Mol. Gen. Genet. 184*:249-254 (1981), both of which are herein incorporated by reference in their entirety).

Synechocystis has been used to study gene regulation by gene replacement through homologous recombination or by gene disruption using antibiotic resistance markers (Pakrasi *et al., EMBO 7*:325-332 (1988), herein incorporated by reference in its entirety). In such gene regulation studies, double reciprocal homologous regions of the host genome flanking the gene of interest recombine to stably integrate the gene of interest into the genome. The gene of interest can be expressed once that gene has been stably integrated into the genome. Biochemical analysis can be performed to study the effect of the replaced or deleted gene.

It is understood that the agents of the present invention may be employed in a *Synechocystis* system.

Exogenous genetic material may be transferred into a plant cell and the plant cell regenerated into a whole, fertile or sterile plant. Exogenous genetic material is any genetic material, whether naturally occurring or otherwise, from any source that is capable of being inserted into any organism. Such genetic material may be transferred into either monocotyledons and dicotyledons including but not limited to the crops, maize and soybean (*See specifically,* Chistou, *Particle Bombardment for Genetic Engineering of Plants,* pp. 63-69 (maize), pp50-60 (soybean), Biotechnology Intelligence Unit. Academic Press, San Diego, California (1996), the entirety of which is herein incorporated by reference and generally Chistou, *Particle Bombardment for Genetic Engineering of Plants,* Biotechnology Intelligence Unit. Academic Press, San Diego, California (1996), the entirety of which is herein incorporated by reference).

Transfer of a nucleic acid that encodes for a protein can result in overexpression of that protein in a transformed cell or transgenic plant. One or more of the proteins or fragments thereof encoded by nucleic acid molecules of the present invention may be overexpressed in a transformed cell or transformed plant. Such overexpression may be the result of transient or stable transfer of the exogenous material.

Exogenous genetic material may be transferred into a plant cell by the use of a DNA vector or construct designed for such a purpose. Design of such a vector is generally within the skill of the art (*See*, Plant Molecular Biology: A Laboratory Manual eds. Clark, Springer, New York (1997), the entirety of which is herein incorporated by reference).

A construct or vector may include a plant promoter to express the protein or protein fragment of choice. A number of promoters which are active in plant cells have been described in the literature. These include the nopaline synthase (NOS) promoter (Ebert *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 84*:5745-5749 (1987), the entirety of which is herein incorporated by reference), the octopine synthase (OCS) promoter (which are carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the caulimovirus promoters such as the cauliflower mosaic virus (CaMV) 19S promoter (Lawton *et al.*, *Plant Mol. Biol. 9*:315-324 (1987), the entirety of which is herein incorporated by reference) and the CAMV 35S promoter (Odell *et al.*, *Nature 313*:810-

812 (1985), the entirety of which is herein incorporated by reference), the figwort mosaic virus 35S-promoter, the light-inducible promoter from the small subunit of ribulose-1,5-bis-phosphate carboxylase (ssRUBISCO), the Adh promoter (Walker *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 84*:6624-6628 (1987), the entirety of which is herein incorporated by reference), the sucrose synthase promoter (Yang *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 87*:4144-4148 (1990), the entirety of which is herein incorporated by reference), the R gene complex promoter (Chandler *et al.*, *The Plant Cell 1*:1175-1183 (1989), the entirety of which is herein incorporated by reference), and the chlorophyll a/b binding protein gene promoter, *etc.* These promoters have been used to create DNA constructs which have been expressed in plants; *see*, *e.g.*, PCT publication WO 84/02913, herein incorporated by reference in its entirety.

Promoters which are known or are found to cause transcription of DNA in plant cells can be used in the present invention. Such promoters may be obtained from a variety of sources such as plants and plant viruses. It is preferred that the particular promoter selected should be capable of causing sufficient expression to result in the production of an effective amount of a protein to cause the desired phenotype. In addition to promoters which are known to cause transcription of DNA in plant cells, other promoters may be identified for use in the current invention by screening a plant cDNA library for genes which are selectively or preferably expressed in the target tissues or cells.

For the purpose of expression in source tissues of the plant, such as the leaf, seed, root or stem, it is preferred that the promoters utilized in the present invention have relatively high expression in these specific tissues. For this purpose, one may choose from a number of promoters for genes with tissue- or cell-specific or -enhanced expression. Examples of such promoters reported in the literature include the chloroplast glutamine synthetase GS2 promoter from pea (Edwards *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 87*:3459-3463 (1990), herein incorporated by reference in its entirety), the chloroplast fructose-1,6-biphosphatase (FBPase) promoter from wheat (Lloyd *et al.*, *Mol. Gen. Genet. 225*:209-216 (1991), herein incorporated by reference in its entirety), the nuclear photosynthetic ST-LS1 promoter from potato (Stockhaus *et*

al., *EMBO J. 8*:2445-2451 (1989), herein incorporated by reference in its entirety), the phenylalanine ammonia-lyase (PAL) promoter and the chalcone synthase (CHS) promoter from *Arabidopsis thaliana*. Also reported to be active in photosynthetically active tissues are the ribulose-1,5-bisphosphate carboxylase (RbcS) promoter from eastern larch (*Larix laricina*), the promoter for the *cab* gene, cab6, from pine (Yamamoto *et al.*, *Plant Cell Physiol. 35*:773-778 (1994), herein incorporated by reference in its entirety), the promoter for the Cab-1 gene from wheat (Fejes *et al.*, *Plant Mol. Biol. 15*:921-932 (1990), herein incorporated by reference in its entirety), the promoter for the CAB-1 gene from spinach (Lubberstedt *et al.*, *Plant Physiol. 104*:997-1006 (1994), herein incorporated by reference in its entirety), the promoter for the cab1R gene from rice (Luan *et al.*, *Plant Cell. 4*:971-981 (1992), the entirety of which is herein incorporated by reference), the pyruvate, orthophosphate dikinase (PPDK) promoter from maize (Matsuoka *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 90*: 9586-9590 (1993), herein incorporated by reference in its entirety), the promoter for the tobacco Lhcb1*2 gene (Cerdan *et al.*, *Plant Mol. Biol. 33*: 245-255. (1997), herein incorporated by reference in its entirety), the *Arabidopsis thaliana* SUC2 sucrose-H+ symporter promoter (Truernit *et al.*, *Planta. 196*: 564-570 (1995), herein incorporated by reference in its entirety), and the promoter for the thylacoid membrane proteins from spinach (psaD, psaF, psaE, PC, FNR, atpC, atpD, cab, rbcS). Other promoters for the chlorophyll a/b-binding proteins may also be utilized in the present invention, such as the promoters for LhcB gene and PsbP gene from white mustard (*Sinapis alba*; Kretsch *et al.*, *Plant Mol. Biol. 28*: 219-229 (1995), the entirety of which is herein incorporated by reference).

For the purpose of expression in sink tissues of the plant, such as the tuber of the potato plant, the fruit of tomato, or the seed of maize, wheat, rice, and barley, it is preferred that the promoters utilized in the present invention have relatively high expression in these specific tissues. A number of promoters for genes with tuber-specific or -enhanced expression are known, including the class I patatin promoter (Bevan *et al.*, *EMBO J. 8*: 1899-1906 (1986); Jefferson *et al.*, *Plant Mol. Biol. 14*: 995-1006 (1990), both of which are herein incorporated by reference in its entirety), the promoter for the potato tuber ADPGPP genes, both the large and

small subunits, the sucrose synthase promoter (Salanoubat and Belliard, *Gene. 60*: 47-56 (1987), Salanoubat and Belliard, *Gene. 84*: 181-185 (1989), both of which are incorporated by reference in their entirety), the promoter for the major tuber proteins including the 22 kd protein complexes and proteinase inhibitors (Hannapel, *Plant Physiol. 101*: 703-704 (1993), herein incorporated by reference in its entirety), the promoter for the granule bound starch synthase gene (GBSS) (Visser *et al.*, *Plant Mol. Biol. 17*: 691-699 (1991), herein incorporated by reference in its entirety), and other class I and II patatins promoters (Koster-Topfer *et al.*, *Mol Gen Genet. 219*: 390-396 (1989); Mignery *et al.*, *Gene. 62*: 27-44 (1988), both of which are herein incorporated by reference in their entirety).

Other promoters can also be used to express a fructose 1,6 bisphosphate aldolase gene in specific tissues, such as seeds or fruits. The promoter for β-conglycinin (Chen *et al.*, *Dev. Genet. 10*: 112-122 (1989), herein incorporated by reference in its entirety) or other seed-specific promoters such as the napin and phaseolin promoters, can be used. The zeins are a group of storage proteins found in maize endosperm. Genomic clones for zein genes have been isolated (Pedersen *et al.*, *Cell 29*: 1015-1026 (1982), herein incorporated by reference in its entirety), and the promoters from these clones, including the 15 kD, 16 kD, 19 kD, 22 kD, 27 kD, and gamma genes, could also be used. Other promoters known to function, for example, in maize, include the promoters for the following genes: *waxy, Brittle, Shrunken* 2, Branching enzymes I and II, starch synthases, debranching enzymes, oleosins, glutelins, and sucrose synthases. A particularly preferred promoter for maize endosperm expression is the promoter for the glutelin gene from rice, more particularly the Osgt-1 promoter (Zheng *et al.*, *Mol. Cell Biol. 13*: 5829-5842 (1993), herein incorporated by reference in its entirety). Examples of promoters suitable for expression in wheat include those promoters for the ADPglucose pyrophosphorylase (ADPGPP) subunits, the granule bound and other starch synthases, the branching and debranching enzymes, the embryogenesis-abundant proteins, the gliadins, and the glutenins. Examples of such promoters in rice include those promoters for the ADPGPP subunits, the granule bound and other starch synthases, the branching enzymes, the debranching enzymes, sucrose synthases, and the glutelins.

A particularly preferred promoter is the promoter for rice glutelin, Osgt-1. Examples of such promoters for barley include those for the ADPGPP subunits, the granule bound and other starch synthases, the branching enzymes, the debranching enzymes, sucrose synthases, the hordeins, the embryo globulins, and the aleurone specific proteins.

Root specific promoters may also be used. An example of such a promoter is the promoter for the acid chitinase gene (Samac *et al.*, *Plant Mol. Biol. 25*: 587-596 (1994), the entirety of which is herein incorporated by reference). Expression in root tissue could also be accomplished by utilizing the root specific subdomains of the CaMV35S promoter that have been identified (Lam *et al.*, *Proc. Natl. Acad. Sci. (U.S.A.) 86*:7890-7894 (1989), herein incorporated by reference in its entirety). Other root cell specific promoters include those reported by Conkling *et al.* (Conkling *et al.*, *Plant Physiol. 93*:1203-1211 (1990), the entirety of which is herein incorporated by reference).

Additional promoters that may be utilized are described, for example, in U.S. Patent Nos. 5,378,619, 5,391,725, 5,428,147, 5,447,858, 5,608,144, 5,608,144, 5,614,399, 5,633,441, 5,633,435, and 4,633,436, all of which are herein incorporated in their entirety. In addition, a tissue specific enhancer may be used (Fromm *et al.*, *The Plant Cell 1*:977-984 (1989), the entirety of which is herein incorporated by reference).

Constructs or vectors may also include, with the coding region of interest, a nucleic acid sequence that acts, in whole or in part, to terminate transcription of that region. For example, such sequences have been isolated including the Tr7 3' sequence and the nos 3' sequence (Ingelbrecht *et al.*, *The Plant Cell 1*:671-680 (1989), the entirety of which is herein incorporated by reference; Bevan *et al.*, *Nucleic Acids Res. 11*:369-385 (1983), the entirety of which is herein incorporated by reference), or the like.

A vector or construct may also include regulatory elements. Examples of such include the Adh intron 1 (Callis *et al.*, *Genes and Develop. 1*:1183-1200 (1987), the entirety of which is herein incorporated by reference), the sucrose synthase intron (Vasil *et al.*, Plant Physiol. 91:1575-1579 (1989), the entirety of which is herein incorporated by reference) and the TMV

omega element (Gallie *et al.*, *The Plant Cell 1*:301-311 (1989), the entirety of which is herein incorporated by reference). These and other regulatory elements may be included when appropriate.

A vector or construct may also include a selectable marker. Selectable markers may also be used to select for plants or plant cells that contain the exogenous genetic material. Examples of such include, but are not limited to, a neo gene (Potrykus *et al.*, *Mol. Gen. Genet. 199*:183-188 (1985), the entirety of which is herein incorporated by reference) which codes for kanamycin resistance and can be selected for using kanamycin, G418, *etc.*; a bar gene which codes for bialaphos resistance; a mutant EPSP synthase gene (Hinchee *et al.*, *Bio/Technology 6*:915-922 (1988), the entirety of which is herein incorporated by reference) which encodes glyphosate resistance; a nitrilase gene which confers resistance to bromoxynil (Stalker *et al.*, *J. Biol. Chem. 263*:6310-6314 (1988), the entirety of which is herein incorporated by reference); a mutant acetolactate synthase gene (ALS) which confers imidazolinone or sulphonylurea resistance (European Patent Application 154,204 (Sept. 11, 1985), the entirety of which is herein incorporated by reference); and a methotrexate resistant DHFR gene (Thillet *et al.*, *J. Biol. Chem. 263*:12500-12508 (1988), the entirety of which is herein incorporated by reference).

A vector or construct may also include a transit peptide. Incorporation of a suitable chloroplast transit peptide may also be employed (European Patent Application Publication Number 0218571, the entirety of which is herein incorporated by reference). Translational enhancers may also be incorporated as part of the vector DNA. DNA constructs could contain one or more 5' non-translated leader sequences which may serve to enhance expression of the gene products from the resulting mRNA transcripts. Such sequences may be derived from the promoter selected to express the gene or can be specifically modified to increase translation of the mRNA. Such regions may also be obtained from viral RNAs, from suitable eukaryotic genes, or from a synthetic gene sequence. For a review of optimizing expression of transgenes, see Koziel *et al.*, *Plant Mol. Biol. 32*:393-405 (1996), the entirety of which is herein incorporated by reference.

A vector or construct may also include a screenable marker. Screenable markers may be used to monitor expression. Exemplary screenable markers include a β-glucuronidase or uidA gene (GUS) which encodes an enzyme for which various chromogenic substrates are known (Jefferson, *Plant Mol. Biol. Rep. 5:* 387-405 (1987), the entirety of which is herein incorporated by reference; Jefferson *et al., EMBO J. 6:* 3901-3907 (1987), the entirety of which is herein incorporated by reference); an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues ((Dellaporta *et al.*, Stadler Symposium *11*:263-282 (1988), the entirety of which is herein incorporated by reference); a β-lactamase gene (Sutcliffe *et al., Proc. Natl. Acad. Sci. (U.S.A.) 75*: 3737-3741 (1978), the entirety of which is herein incorporated by reference), a gene which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a luciferase gene (Ow *et al., Science 234*: 856-859 (1986), the entirety of which is herein incorporated by reference) a xylE gene (Zukowsky *et al., Proc. Natl. Acad. Sci. (U.S.A.) 80*:1101-1105 (1983), the entirety of which is herein incorporated by reference) which encodes a catechol dioxygenase that can convert chromogenic catechols; an α-amylase gene (Ikatu *et al., Bio/Technol. 8*:241-242 (1990), the entirety of which is herein incorporated by reference); a tyrosinase gene (Katz *et al., J. Gen. Microbiol. 129*:2703-2714 (1983), the entirety of which is herein incorporated by reference) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to melanin; an α-galactosidase, which will turn a chromogenic α-galactose substrate.

Included within the terms "selectable or screenable marker genes" are also genes which encode a scriptable marker whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes which can be detected catalytically. Secretable proteins fall into a number of classes, including small, diffusible proteins detectable, *e.g.*, by ELISA, small active enzymes detectable in extracellular solution (*e.g.*, α-amylase, β-lactamase, phosphinothricin transferase), or proteins which are inserted or

trapped in the cell wall (such as proteins which include a leader sequence such as that found in the expression unit of extension or tobacco PR-S). Other possible selectable and/or screenable marker genes will be apparent to those of skill in the art.

Methods and compositions for transforming a bacteria and other microorganisms are known in the art (see for example Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1989), the entirety of which is herein incorporated by reference).

There are many methods for introducing transforming nucleic acid molecules into plant cells. Suitable methods are believed to include virtually any method by which nucleic acid molecules may be introduced into a cell, such as by *Agrobacterium* infection or direct delivery of nucleic acid molecules such as, for example, by PEG-mediated transformation, by electroporation or by acceleration of DNA coated particles, *etc.* (Pottykus, *Ann. Rev. Plant Physiol. Plant Mol. Biol. 42*:205-225 (1991), the entirety of which is herein incorporated by reference; Vasil, *Plant Mol. Biol. 25*: 925-937 (1994), the entirety of which is herein incorporated by reference. For example, electroporation has been used to transform maize protoplasts (Fromm *et al.*, *Nature 312*:791-793 (1986), the entirety of which is herein incorporated by reference).

Other vector systems suitable for introducing transforming DNA into a host plant cell includes but is not limited to binary artificial chromosome (BIBAC) vectors (Hamilton *et al.*, *Gene 200*:107-116, (1997), the entirety of which is herein incorporated by reference, and transfection with RNA viral vectors (Della-Cioppa *et al.*, *Ann. N.Y. Acad. Sci.* (1996), 792 (Engineering Plants for Commercial Products and Applications), 57-61, the entirety of which is herein incorporated by reference.

Technology for introduction of DNA into cells is well known to those of skill in the art. Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, *Virology, 54*:536-539 (1973), the entirety of which is herein incorporated by reference); (2) physical methods such as microinjection (Capecchi, *Cell 22*:479-

488 (1980), electroporation (Wong and Neumann, *Biochem. Biophys. Res. Commun., 107*:584-587 (1982); Fromm *et al., Proc. Natl. Acad. Sci. USA*, 82:5824-5828 (1985); U. S. Patent No. 5,384,253; and the gene gun (Johnston and Tang, *Methods Cell Biol. 43*:353-365 (1994), all of which the entirety is herein incorporated by reference; (3) viral vectors (Clapp, *Clin. Perinatol., 20*:155-168 (1993); Lu *et al., J. Exp. Med., 178*:2089-2096 (1993); Eglitis and Anderson, *Biotechniques, 6*:608-614 (1988), all of which the entirety is herein incorporated by reference); and (4) receptor-mediated mechanisms (Curiel *et al., Hum. Gen. Ther., 3*:147-154 (1992); Wagner *et al., Proc. Natl. Acad. Sci. USA, 89*:6099-6103 (1992), all of which the entirety is herein incorporated by reference).

Acceleration methods that may be used include, for example, microprojectile bombardment and the like. One example of a method for delivering transforming nucleic acid molecules to plant cells is microprojectile bombardment. This method has been reviewed by Yang and Christou, eds., *Particle Bombardment Technology for Gene Transfer*, Oxford Press, Oxford, England (1994), the entirety of which is herein incorporated by reference). Non-biological particles (microprojectiles) that may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum, and the like.

A particular advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly, and stably transforming monocotyledons, is that neither the isolation of protoplasts (Cristou *et al., Plant Physiol. 87*:671-674 (1988), the entirety of which is herein incorporated by reference) nor the susceptibility of *Agrobacterium* infection is required. An illustrative embodiment of a method for delivering DNA into maize cells by acceleration is a biolistics g-particle delivery system, which can be used to propel particles coated with DNA through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with corn cells cultured in suspension. Gordon-Kamm *et al.*, describes the basic procedure for coating tungsten particles with DNA (Gordon-Kamm *et al., Plant Cell 2*: 603-618 (1990), the entirety of which is herein incorporated by reference). The screen disperses the tungsten nucleic acid

particles so that they are not delivered to the recipient cells in large aggregates. A particle delivery system suitable for use with the present invention is the helium acceleration PDS-1000/He gun which is available from Bio-Rad Laboratories (Bio-Rad, Hercules, California)(Sanford *et al.*, *Technique 3*:3-16 (1991), the entirety of which is herein incorporated by reference).

For the bombardment, cells in suspension may be concentrated on filters. Filters containing the cells to be bombarded are positioned at an appropriate distance below the microprojectile stopping plate. If desired, one or more screens are also positioned between the gun and the cells to be bombarded.

Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of techniques set forth herein one may obtain up to 1000 or more foci of cells transiently expressing a marker gene. The number of cells in a focus which express the exogenous gene product 48 hours post-bombardment often range from one to ten and average one to three.

In bombardment transformation, one may optimize the prebombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos. In another alternative embodiment, plastids can be stably transformed. Methods disclosed for plastid transformation in

higher plants include the particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination (Svab *et al., Proc. Natl. Acad. Sci. (U.S.A.) 87*:8526-8530 (1990); Svab and Maliga, *Proc. Natl. Acad. Sci. (U.S.A.) 90*:913-917 (1993); Staub and Maliga, *EMBO J. 12*:601-606 (1993); U.S. Patents 5, 451,513 and 5,545,818, all of which are herein incorporated by reference in their entirety).

Accordingly, it is contemplated that one may wish to adjust various aspects of the bombardment parameters in small scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance, and helium pressure. One may also minimize the trauma reduction factors by modifying conditions which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example the methods described (Fraley *et al., Biotechnology 3*:629-635 (1985); Rogers *et al., Meth. In Enzymol, 153*:253-277 (1987), both of which are herein incorporated by reference in their entirety. Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al., Mol. Gen. Genet., 205*:34 (1986), the entirety of which is herein incorporated by reference).

Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al., In: Plant DNA Infectious Agents*, T. Hohn and J. Schell, eds., Springer-Verlag, New York, pp. 179-

203 (1985), the entirety of which is herein incorporated by reference. Moreover, recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes (Rogers *et al., Meth. In Enzymol., 153*:253-277 (1987), the entirety of which is herein incorporated by reference). In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single gene on one chromosome. Such transgenic plants can be referred to as being heterozygous for the added gene. More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.*, a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for the gene of interest.

It is also to be understood that two different transgenic plants can also be mated to produce offspring that contain two independently segregating added, exogenous genes. Selfing of appropriate progeny can produce plants that are homozygous for both added, exogenous genes that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated, as is vegetative propagation.

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments. See for example (Potrykus *et al., Mol. Gen. Genet., 205*:193-200 (1986); Lorz *et al., Mol. Gen. Genet., 199*:178, (1985); Fromm *et al., Nature, 319*:791,(1986); Uchimiya *et al.*,

Mol. Gen. Genet.:204:204, (1986); Callis *et al., Genes and Development, 1183*,(1987); Marcotte *et al., Nature, 335*:454, (1988), all of which the entirety is herein incorporated by reference).

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts are described (Fujimura *et al., Plant Tissue Culture Letters, 2*:74,(1985); Toriyama *et al., Theor Appl. Genet. 205*:34. (1986); Yamada *et al., Plant Cell Rep., 4*:85, (1986); Abdullah *et al., Biotechnology, 4*:1087, (1986), all of which the entirety is herein incorporated by reference).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, *Biotechnology, 6*:397,(1988), the entirety of which is herein incorporated by reference). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil *et al., Bio/Technology 10*:667, (1992), the entirety of which is herein incorporated by reference).

Using the latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al., Nature, 328*:70, (1987); Klein *et al., Proc. Natl. Acad. Sci. USA*, 85:8502-8505, (1988); McCabe *et al., Biotechnology, 6*:923, (1988), all of which the entirety is herein incorporated by reference). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

Other methods of cell transformation can also be used and include but are not limited to introduction of DNA into plants by direct DNA transfer into pollen (Zhou *et al., Methods in Enzymology, 101*:433, (1983); Hess *et al., Intern Rev. Cytol., 107*:367, (1987); Luo *et al., Plant Mol Biol. Reporter, 6*:165, (1988), all of which the entirety is herein incorporated by reference), by direct injection of DNA into reproductive organs of a plant (Pena *et al.*, Nature, 325:274, (1987), the entirety of which is herein incorporated by reference), or by direct injection of DNA into the cells of immature embryos followed by the rehydration of dessicated embryos (Neuhaus

et al., Theor. Appl. Genet., 75:30, (1987), the entirety of which is herein incorporated by reference).

The regeneration, development, and cultivation of plants from single plant protoplast transformants or from various transformed explants is well known in the art (Weissbach and Weissbach, *In: Methods for Plant Molecular Biology*, (Eds.), Academic Press, Inc. San Diego, CA, (1988), the entirety of which is herein incorporated by reference). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a protein of interest is well known in the art. Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants, as discussed before. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important lines. Conversely, pollen from plants of these important lines is used to pollinate regenerated plants. A transgenic plant of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

There are a variety of methods for the regeneration of plants from plant tissue. The particular method of regeneration will depend on the starting plant tissue and the particular plant species to be regenerated.

Methods for transforming dicots, primarily by use of *Agrobacterium tumefaciens*, and obtaining transgenic plants have been published for cotton (U. S. Patent No. 5,004,863, U.S. Patent No. 5,159,135, U.S. Patent No. 5,518,908, all of which the entirety is herein incorporated by reference); soybean (U. S. Patent No. 5,569,834, U. S. Patent No. 5,416,011, McCabe *et al.*, *Biotechnology 6*:923, (1988), Christou *et al., Plant Physiol., 87*:671-674 (1988), all of which the entirety is herein incorporated by reference); *Brassica* (U. S. Patent No. 5,463,174, the entirety of which is herein incorporated by reference); peanut (Cheng *et al., Plant Cell Rep. 15*: 653-657

(1996), McKently *et al.*, *Plant Cell Rep. 14*:699-703 (1995), all of which the entirety is herein incorporated by reference); papaya (Yang *et al.*, (1996), the entirety of which is herein incorporated by reference); pea (Grant *et al.*, *Plant Cell Rep. 15*:254-258, (1995), the entirety of which is herein incorporated by reference).

Transformation of monocotyledons using electroporation, particle bombardment, and *Agrobacterium* have also been reported. Transformation and plant regeneration have been achieved in asparagus (Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA 84*:5345, (1987), the entirety of which is herein incorporated by reference); barley (Wan and Lemaux, *Plant Physiol 104*:37, (1994), the entirety of which is herein incorporated by reference); maize (Rhodes *et al.*, *Science 240*: 204, (1988), Gordon-Kamm *et al.*, *Plant Cell, 2*:603, (1990), Fromm *et al.*, *Bio/Technology 8*:833, (1990), Koziel *et al.*, *Bio/Technology 11*:194, (1993), Armstrong *et al.*, *Crop Science 35*:550-557, (1995), all of which the entirety is herein incorporated by reference); oat (Somers *et al.*, *Bio/Technology, 10*:1589, (1992), the entirety of which is herein incorporated by reference); orchardgrass (Horn *et al.*, *Plant Cell Rep. 7*:469, (1988), the entirety of which is herein incorporated by reference); rice (Toriyama *et al.*, *Theor Appl. Genet. 205*:34, (1986); Park *et al.*, *Plant Mol. Biol.,32*: 1135-1148, (1996); Abedinia *et al.*, *Aust. J. Plant Physiol.24*:133-141, (1997); Zhang and Wu, *Theor. Appl. Genet. 76*:835, (1988); Zhang *et al. Plant Cell Rep. 7*:379, (1988); Battraw and Hall, *Plant Sci. 86*:191-202, (1992); Christou *et al.*, *Bio/Technology 9*:957, (1991), all of which the entirety is herein incorporated by reference); sugarcane (Bower and Birch, Plant J. 2:409, (1992), the entirety of which is herein incorporated by reference); tall fescue (Wang *et al.*, Bio/Technology 10:691, (1992), the entirety of which is herein incorporated by reference), and wheat (Vasil *et al.*, Bio/Technology 10:667, (1992), the entirety of which is herein incorporated by reference; U. S. Patent No. 5,631,152, the entirety of which is herein incorporated by reference.

Assays for gene expression based on the transient expression of cloned nucleic acid constructs have been developed by introducing the nucleic acid molecules into plant cells by polyethylene glycol treatment, electroporation, or particle bombardment (Marcotte, *et al.*, *Nature*,

335: 454-457 (1988), the entirety of which is herein incorporated by reference; Marcotte, *et al.*, *Plant Cell, 1:* 523-532 (1989), the entirety of which is herein incorporated by reference; McCarty, *et al., Cell 66:* 895-905 (1991), the entirety of which is herein incorporated by reference; Hattori, *et al., Genes Dev. 6:* 609-618 (1992), the entirety of which is herein incorporated by reference; Goff, *et al., EMBO J. 9:* 2517-2522 (1990), the entirety of which is herein incorporated by reference). Transient expression systems may be used to functionally dissect gene constructs (*See generally*, Mailga *et al., Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995)).

Any of the nucleic acid molecules of the present invention may be introduced into a plant cell in a permanent or transient manner in combination with other genetic elements such as vectors, promoters enhancers *etc.* Further any of the nucleic acid molecules of the present invention may be introduced into a plant cell in a manner that allows for over expression of the protein or fragment thereof encoded by the nucleic acid molecule.

Cosuppression is the reduction in expression levels, usually at the level of RNA, of a particular endogenous gene or gene family by the expression of a homologous sense construct that is capable of transcribing mRNA of the same strandedness as the transcript of the endogenous gene (Napoli *et al., Plant Cell 2*: 279-289 (1990), the entirety of which is herein incorporated by reference; van der Krol *et al., Plant Cell 2*: 291-299 (1990), the entirety of which is herein incorporated by reference). Cosuppression may result from stable transformation with a single copy nucleic acid molecule that is homologous to a nucleic acid sequence found with the cell (Prolls and Meyer, *Plant J. 2*:465-475 (1992), the entirety of which is herein incorporated by reference) or with multiple copies of a nucleic acid molecule that is homologous to a nucleic acid sequence found with the cell (Mittlesten *et al., Mol. Gen. Genet. 244*: 325-330 (1994), the entirety of which is herein incorporated by reference). Genes, even though different, linked to homologous promoters may result in the cosuppression of the linked genes (Vaucheret, *C.R. Acad. Sci. III 316*: 1471-1483 (1993), the entirety of which is herein incorporated by reference).

This technique has, for example been applied to generate white flowers from red petunia and tomatoes that do not ripen on the vine. Up to 50% of petunia transformants that contained a sense copy of the chalcone synthase (CHS) gene produced white flowers or floral sectors; this was as a result of the post-transcriptional loss of mRNA encoding CHS (Flavell, *Proc. Natl. Acad. Sci. (U.S.A.) 91*:3490-3496 (1994)), the entirety of which is herein incorporated by reference). Cosuppression may require the coordinate transcription of the transgene and the endogenous gene, and can be reset by a developmental control mechanism (Jorgensen, *Trends Biotechnol, 8*:340344 (1990), the entirety of which is herein incorporated by reference; Meins and Kunz, In: *Gene Inactivation and Homologous Recombination in Plants* (Paszkowski, J., ed.), pp. 335-348. Kluwer Academic, Netherlands (1994), the entirety of which is herein incorporated by reference).

It is understood that one or more of the nucleic acids of the present invention including those comprising SEQ ID NO:1 through SEQ ID NO:32236 or complement thereof or fragments of either or other nucleic acid molecules of the present invention may be introduced into a plant cell and transcribed using an appropriate promoter with such transcription resulting in the co-suppression of an endogenous protein.

Antisense approaches are a way of preventing or reducing gene function by targeting the genetic material (Mol *et al., FEBS Lett. 268*: 427-430 (1990), the entirety of which is herein incorporated by reference). The objective of the antisense approach is to use a sequence complementary to the target gene to block its expression and create a mutant cell line or organism in which the level of a single chosen protein is selectively reduced or abolished. Antisense techniques have several advantages over other 'reverse genetic' approaches. The site of inactivation and its developmental effect can be manipulated by the choice of promoter for antisense genes or by the timing of external application or microinjection. Antisense can manipulate its specificity by selecting either unique regions of the target gene or regions where it shares homology to other related genes (Hiatt *et al., In Genetic Engineering*, Setlow (ed.), Vol. 11, New York: Plenum 49-63 (1989), the entirety of which is herein incorporated by reference).

The principle of regulation by antisense RNA is that RNA that is complementary to the target mRNA is introduced into cells, resulting in specific RNA:RNA duplexes being formed by base pairing between the antisense substrate and the target mRNA (Green *et al.*, *Annu. Rev. Biochem. 55*: 569-597 (1986), the entirety of which is herein incorporated by reference). Under one embodiment, the process involves the introduction and expression of an antisense gene sequence. Such a sequence is one in which part or all of the normal gene sequences are placed under a promoter in inverted orientation so that the 'wrong' or complementary strand is transcribed into a noncoding antisense RNA that hybridizes with the target mRNA and interferes with its expression (Takayama and Inouye, *Crit. Rev. Biochem. Mol. Biol. 25*: 155-184 (1990), the entirety of which is herein incorporated by reference). An antisense vector is constructed by standard procedures and introduced into cells by transformation, transfection, electroporation, microinjection, or by infection, *etc.* The type of transformation and choice of vector will determine whether expression is transient or stable. The promoter used for the antisense gene may influence the level, timing, tissue, specificity, or inducibility of the antisense inhibition.

It is understood that protein synthesis activity in a plant cell may be reduced or depressed by growing a transformed plant cell containing a nucleic acid molecule whose non-transcribed strand encodes a protein or fragment thereof.

Antibodies have been expressed in plants (Hiatt *et al.*, *Nature 342:*76-78 (1989), the entirety of which is herein incorporated by reference; Conrad and Fielder, *Plant Mol. Biol. 26:* 1023-1030 (1994), the entirety of which is herein incorporated by reference). cytoplasmic expression of a scFv (single-chain Fv antibodies) has been reported to delay infection by artichoke mottled crinkle virus. Transgenic plants that express antibodies directed against endogenous proteins may exhibit a physiological effect (Philips *et al.*, *EMBO J. 16:* 4489-4496 (1997), the entirety of which is herein incorporated by reference; Marion-Poll, *Trends in Plant Science 2:* 447-448 (1997), the entirety of which is herein incorporated by reference). For example, expressed anti-abscisic antibodies reportedly result in a general perturbation of seed development (Philips *et al.*, *EMBO J. 16:* 4489-4496 (1997)).

Antibodies that are catalytic may also be expressed in plants (abzymes). The principle behind abzymes is that since antibodies may be raised against many molecules, this recognition ability can be directed toward generating antibodies that bind transition states to force a chemical reaction forward (Persidas, *Nature Biotechnology 15*:1313-1315 (1997), the entirety of which is herein incorporated by reference; Baca *et al.*, *Ann. Rev. Biophys. Biomol. Struct. 26*:461-493 (1997), the entirety of which is herein incorporated by reference). The catalytic abilities of abzymes may be enhanced by site directed mutagenesis. Examples of abzymes are, for example, set forth in U.S. Patent No: 5,658,753; U.S. Patent No. 5,632,990; U.S. Patent No. 5,631,137; U.S. Patent 5,602,015; U.S. Patent No. 5,559,538; U.S. Patent No. 5,576,174; U.S. Patent No. 5,500,358; U.S. Patent 5,318,897; U.S. Patent No. 5,298,409; U.S. Patent No. 5,258,289 and U.S. Patent No. 5,194,585, all of which are herein incorporated in their entirety.

It is understood that any of the antibodies of the present invention may be expressed in plants and that such expression can result in a physiological effect. It is also understood that any of the expressed antibodies may be catalytic.

In addition to the above discussed procedures, practitioners are familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (*e.g.*, DNA molecules, plasmids, *etc.*), generation of recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al., Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Mailga *et al., Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995), the entirety of which is herein incorporated by reference; Birren *et al., Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, New York, the entirety of which is herein incorporated by reference).

The nucleotide sequence provided in SEQ ID NO:1, through SEQ ID NO:32236 or fragment thereof, or complement thereof, or a nucleotide sequence at least 90% identical, preferably 95%, identical even more preferably 99% or 100% identical to the sequence provided in SEQ ID NO:1 through SEQ ID NO:32236 or fragment thereof, or complement thereof, can be

"provided" in a variety of mediums to facilitate use fragment thereof. Such a medium can also provide a subset thereof in a form that allows a skilled artisan to examine the sequences.

In one application of this embodiment, a nucleotide sequence of the present invention can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc, storage medium, and magnetic tape: optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. A skilled artisan can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising computer readable medium having recorded thereon a nucleotide sequence of the present invention.

As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate media comprising the nucleotide sequence information of the present invention. A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing one or more of nucleotide sequences of the present invention, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul *et al., J. Mol. Biol. 215:*403-410 (1990)) and BLAZE (Brutlag *et al., Comp. Chem. 17:*203-207 (1993), the entirety of which is herein incorporated by reference) search algorithms on a Sybase system can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs or proteins from other organisms. Such ORFs are protein-encoding fragments within the sequences of the present invention and are useful in producing commercially important proteins such as enzymes used in amino acid biosynthesis, metabolism, transcription, translation, RNA processing, nucleic acid and a protein degradation, protein modification, and DNA replication, restriction, modification, recombination, and repair.

The present invention further provides systems, particularly computer-based systems, which contain the sequence information described herein. Such systems are designed to identify commercially important fragments of the nucleic acid molecule of the present invention. As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention.

As indicated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory that can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures

having recorded thereon the nucleotide sequence information of the present invention. As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequence of the present invention that match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are available and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTIN and BLASTIX (NCBIA). One of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems.

The most preferred sequence length of a target sequence is from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that during searches for commercially important fragments of the nucleic acid molecules of the present invention, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequences or sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzymatic active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, cis elements, hairpin structures and inducible expression elements (protein binding sequences).

Thus, the present invention further provides an input means for receiving a target sequence, a data storage means for storing the target sequences of the present invention sequence identified using a search means as described above, and an output means for outputting the identified homologous sequences. A variety of structural formats for the input and output means

can be used to input and output information in the computer-based systems of the present invention. A preferred format for an output means ranks fragments of the sequence of the present invention by varying degrees of homology to the target sequence or target motif. Such presentation provides a skilled artisan with a ranking of sequences which contain various amounts of the target sequence or target motif and identifies the degree of homology contained in the identified fragment.

A variety of comparing means can be used to compare a target sequence or target motif with the data storage means to identify sequence fragments sequence of the present invention. For example, implementing software which implement the BLAST and BLAZE algorithms (Altschul *et al., J. Mol. Biol. 215:*403-410 (1990)) can be used to identify open frames within the nucleic acid molecules of the present invention. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer-based systems of the present invention. Having now generally described the invention, the same will be more readily understood through reference to the following examples which are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

Example 1

The normalized cDNA libraries LIB3115 and LIB3116 are generated from maize (RX601, Asgrow Seed Company, Des Moines, Iowa U.S.A.) pooled leaf tissue harvested from field grown plants at Asgrow research stations. Leaves are harvested at antithesis from open pollinated plants in a field (multiple row) setting. The ear leaves from 10-12 plants are harvested, pooled, frozen in liquid nitrogen and then frozen at -80°C where they are stored until RNA preparation. RNA is prepared and libraries are constructed as in Example 2.

The normalized library LIB3115 is prepared from single-stranded plasmid cDNA and the normalized library LIB3116, from double-stranded plasmid cDNA. Single-stranded plasmid cDNA or double-stranded plasmid cDNA representing approximately 1 X 10^6 colony forming units is isolated using standard protocols and used as target. Biotinylated genomic corn DNA is

used as driver. The biotinylated genomic corn DNA is mixed with the single stranded plasmid cDNA or double stranded plasmid cDNA in a 1:1 molar ratio and allowed to hybridize. Genomic DNA-plasmid cDNA hybrids are captured on Dynabeads M280 streptavidin. The dynabeads with captured hybrids are collected with a magnet. Captured hybrids are eluted in water. The resulting clones are subjected to a second round of hybridization identical to the first. SEQ ID NO: 1 through SEQ ID NO: 2177 are from LIB3115. SEQ ID NO: 2178 through SEQ ID NO: 4013 are from LIB3116.

The cDNA libraries LIB3117 and LIB3118 are generated from maize (RX601, Asgrow Seed Company, Des Moines, Iowa U.S.A) pooled kernels from plants at 15 to 20 days after pollination. Samples are collected from field grown material at Asgrow research stations. Whole kernels from hand pollinated (control pollination) plants are harvested as whole ears and immediately frozen on dry ice. Kernels from each of 10-12 ears from each location are pooled and ground together in liquid nitrogen. The harvested tissue is then stored at –80°C until RNA preparation. RNA is prepared and libraries are constructed as in Example 2.

The normalized library LIB3117 is prepared from single-stranded plasmid cDNA and the normalized library LIB3118, from double-stranded plasmid cDNA. Single-stranded plasmid cDNA or double-stranded plasmid cDNA representing approximately 1 X 10^6 colony forming units is isolated using standard protocols and used as target. Biotinylated genomic corn DNA is used as driver. The biotinylated genomic corn DNA is mixed with the single stranded plasmid cDNA or double stranded plasmid cDNA in a 1:1 molar ratio and allowed to hybridize. Genomic DNA-plasmid cDNA hybrids are captured on Dynabeads M280 streptavidin. The dynabeads with captured hybrids are collected with a magnet. Captured hybrids are eluted in water. The resulting clones are subjected to a second round of hybridization identical to the first. SEQ ID NO: 4014 through SEQ ID NO: 4629 are from LIB3117. SEQ ID NO: 4630 through SEQ ID NO: 7156 are from LIB3118.

The cDNA library (LIB3150) is prepared from endosperm harvested from 5-8 days after pollination maize plants which are beyond the V10 stage. *Zea mays* genotype H99 (USDA

Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and about 3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected at 5, 6, 7, and 8 days after pollination and then dissected into embryos and endosperms. After dissection, the endosperm is then immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. RNA preparation and library construction is as in Example 2. SEQ ID NO: 7157 through SEQ ID NO: 16535 are from LIB 3150.

The cDNA library (LIB3151) is prepared from endosperm harvested from 18 days after pollination maize plants which are beyond the V10 stage. *Zea mays* genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and ~3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected at 18 days after pollination and then dissected into embryos and endosperms. After dissection, the endosperm is then immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. RNA is prepared and library

construction is as in Example 2. SEQ ID NO: 16536 through SEQ ID NO: 20787 are from LIB3151.

The cDNA library (LIB3152) is prepared from endosperms harvested from 26 DAP maize plants which are beyond the V10 stage. *Zea mays* genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and ~3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected at 26 days after pollination and then dissected into embryos and endosperms. After dissection, the endosperms are then immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. RNA is prepared and library construction is as in Example 2. SEQ ID NO: 20788 through SEQ ID NO: 24930 are from LIB3152.

The cDNA library (LIB3180) is prepared from kernels harvested from 10 DAP maize plants which are beyond the V10 stage. *Zea mays* genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and ~3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Kernels are collected at 10 days after pollination, immediately frozen in liquid nitrogen and then

stored at -80°C until RNA preparation. RNA is prepared and library construction is as in Example 2. SEQ ID NO: 26366 through SEQ ID NO: 30743 are from LIB3180.

The cDNA library (LIB3152) is prepared from endosperms harvested from 26 DAP maize plants which are beyond the V10 stage. *Zea mays* genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and ~3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected at 26 days after pollination and then dissected into embryos and endosperm. After dissection, the endosperm is then immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. RNA is prepared and library construction is as in Example 2. SEQ ID NO: 20788 through SEQ ID NO: 24930 are from LIB3152.

The subtractive cDNA library LIB3153 is generated by subtracting driver cDNA, which is prepared from kernels harvested from 15 DAP maize plants, from target cDNA, which is prepared from endosperms harvested from 5-8 day after pollination (DAP) maize plants. *Zea mays* genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and about 3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70

F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected at 5 to 8 and 15 days after pollination. The 5 to 8 DAP kernels are dissected to separate endosperms from embryos and the harvested endosperms are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. The 15 DAP kernels are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. The RNA is purified from the stored tissue and the subtractive cDNA library is constructed as described in Example 2. SEQ ID NO: 24931 through SEQ ID NO: 25680 are from LIB3153.

The subtractive cDNA library LIB3154 is generated by subtracting driver cDNA, which is prepared from endosperms harvested from 5 to 8 DAP maize plants, from target cDNA, which is prepared from endosperm harvested from 18 day after pollination (DAP) maize plants. *Zea mays* genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and about 3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected both at 5 to 8 DAP and at 18 DAP and then dissected to separate endosperms from embryos. The harvested 5 to 8 DAP and 18 DAP endosperms are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. The RNA is purified from the stored tissue and the subtractive cDNA library is constructed as described in Example 2. SEQ ID NO: 25681 through SEQ ID NO: 26365 are from LIB3154.

The subtractive cDNA library LIB3181 is generated by subtracting driver cDNA, which is prepared from kernels harvested from 15 DAP maize plants, from target cDNA, which is prepared from endosperms harvested from 26 day after pollination (DAP) maize plants. *Zea*

mays genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and about 3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected at 15 and 26 DAP. The 15 DAP kernels are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. The harvested 26 DAP kernel is dissected to separate endosperms from embryos and the endosperms are immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. The RNA is purified from the stored tissue and the subtractive cDNA library is constructed as described in Example 2. SEQ ID NO: 30744 through SEQ ID NO: LIB31501 are from LIB3181.

The subtractive cDNA library LIB3182 is generated by subtracting driver cDNA, which is prepared from endosperm harvested from 26 DAP maize plants, from target cDNA, which is prepared from endosperm harvested from 18 day after pollination (DAP) maize plants. *Zea mays* genotype H99 (USDA Maize Genetic Stock Center, Urbana, Illinois U.S.A.) seeds are planted at a depth of approximately 3 cm in soil into 2"-3" peat pots containing Metro 200 growing medium. After 2-3 weeks growth, they are transplanted into 10" pots containing the same. Plants are watered daily before transplantation and about 3 times a week after transplantation. Peters 15-16-17 fertilizer is applied three times per week after transplanting at a strength of 150 ppm N. Two to three times during the lifetime of the plant, from transplanting to flowering, a total of 900 mg Fe is added to each pot. Corn plants are grown in a green house in 15hr day /9hr night cycles. The daytime temperature is 80 F and the night time temperature is 70 F. Lighting is provided by 1000 W sodium vapor lamps. Whole kernels are collected both at 18 and 26 DAP

and then dissected to separate endosperm from embryos. The harvested 18 DAP and 26 DAP endosperm is immediately frozen in liquid nitrogen and then stored at -80°C until RNA preparation. The RNA is purified from the stored tissue and the subtractive cDNA library is constructed as described in Example 2. SEQ ID NO: 31502 through SEQ ID NO: LIB32236 are from LIB3182.

Example 2

The stored RNA is purified using Trizol reagent from Life Technologies (Gibco BRL, Life Technologies, Gaithersburg, Maryland U.S.A.), essentially as recommended by the manufacturer. Poly A+ RNA (mRNA) is purified using magnetic oligo dT beads essentially as recommended by the manufacturer (Dynabeads, Dynal Corporation, Lake Success, New York U.S.A.).

Construction of plant cDNA libraries is well-known in the art and a number of cloning strategies exist. A number of cDNA library construction kits are commercially available. The Superscript™ Plasmid System for cDNA synthesis and Plasmid Cloning (Gibco BRL, Life Technologies, Gaithersburg, Maryland U.S.A.) is used, following the conditions suggested by the manufacturer.

Example 3

The cDNA libraries are plated on LB agar containing the appropriate antibiotics for selection and incubated at 37° for a sufficient time to allow the growth of individual colonies. Single selective media colonies are individually placed in each well of a 96-well microtiter plates containing LB liquid including the selective antibiotics. The plates are incubated overnight at approximately 37°C with gentle shaking to promote growth of the cultures. The plasmid DNA is isolated from each clone using Qiaprep plasmid isolation kits, using the conditions recommended by the manufacturer (Qiagen Inc., Santa Clara, California U.S.A.).

Template plasmid DNA clones are used for subsequent sequencing. For sequencing, the ABI PRISM dRhodamine Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq® DNA Polymerase, FS, is used (PE Applied Biosystems, Foster City, California U.S.A.).

We claim:

Sub. a1

1. A substantially purified nucleic acid molecule that encodes a maize protein or fragment thereof comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 4013.

2. A substantially purified maize protein or fragment thereof, wherein said maize protein is encoded by a nucleic acid molecule that comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 4013

3. A transformed plant having a nucleic acid molecule which comprises:

(a) an exogenous promoter region which functions in a plant cell to cause the production of a mRNA molecule;

(b) a structural nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through SEQ ID NO: 4013 or complements thereof;

(c) a 3' non-translated sequence that functions in said plant cell to cause termination of transcription and addition of polyadenylated ribonucleotides to a 3' end of said mRNA molecule.

4. The transformed plant according to claim 3, wherein said structural nucleic acid molecule is a complement of any of the nucleic acid sequences of SEQ ID NO: 1 through SEQ ID NO: 4013.

5. The transformed plant according to claim 4, wherein said plant is maize or soybean.

6. The transformed plant according to claim 4, wherein said plant is maize.

7. The transformed plant according to claim 4, wherein said plant is soybean.

Abstract

The present invention is in the field of plant biochemistry. More specifically the invention relates to nucleic acid molecules that encode proteins and fragments of proteins produced in plant cells, in particular, maize plants. The invention also relates to proteins and fragments of proteins so encoded and antibodies capable of binding the proteins. The invention also relates to methods of using the nucleic acid molecules, proteins and fragments of proteins.